

5

## Compositions and Methods for Recombinational Cloning of Nucleic Acid Molecules

### CROSS REFERENCE TO RELATED APPLICATIONS

The present application claims the benefit of U.S. Provisional Application No. 60/108,324, filed November 13, 1998, the contents of which are incorporated by reference herein in their entirety.

10

### BACKGROUND OF THE INVENTION

#### *Field of the Invention*

15  
The present invention relates generally to recombinant DNA technology. The invention relates more specifically to compositions and methods for recombinational cloning of nucleic acid molecules using recombination systems. In particular, the invention relates to compositions comprising one or more ribosomal proteins, preferably one or more prokaryotic ribosomal proteins and particularly one or more *E. coli* ribosomal proteins, and one or more additional components required for recombinational cloning (such as one or more recombination proteins), and the use of these compositions in methods of recombinational cloning of nucleic acid molecules. The invention also relates to  
20  
isolated nucleic acid molecules produced by the methods of the invention, to vectors comprising such nucleic acid molecules, and to host cells comprising such nucleic acid molecules and vectors.

#### *Related Art*

#### **Site-specific recombinases**

Site-specific recombinases are proteins that are present in many organisms (e.g. viruses and bacteria) and have been characterized to have both endonuclease

25

30

and ligase properties. These recombinases (along with associated proteins in some cases) recognize specific sequences of bases in DNA and exchange the DNA segments flanking those segments. The recombinases and associated proteins are collectively referred to as "recombination proteins" (see, e.g., Landy, A., *Current Opinion in Biotechnology* 3:699-707 (1993)).

Numerous recombination systems from various organisms have been described. See, e.g., Hoess *et al.*, *Nucleic Acids Research* 14(6):2287 (1986); Abremski *et al.*, *J. Biol. Chem.* 261(1):391 (1986); Campbell, *J. Bacteriol.* 174(23):7495 (1992); Qian *et al.*, *J. Biol. Chem.* 267(11):7794 (1992); Araki *et al.*, *J. Mol. Biol.* 225(1):25 (1992); Maeser and Kahnmann *Mol. Gen. Genet.* 230:170-176 (1991); Esposito *et al.*, *Nucl. Acids Res.* 25(18):3605 (1997).

Many of these belong to the integrase family of recombinases (Argos *et al.* *EMBO J.* 5:433-440 (1986)). Perhaps the best studied of these are the Integrase/att system from bacteriophage  $\lambda$  (Landy, A. *Current Opinions in Genetics and Devel.* 3:699-707 (1993)), the Cre/loxP system from bacteriophage P1 (Hoess and Abremski (1990) In *Nucleic Acids and Molecular Biology*, vol. 4. Eds.: Eckstein and Lilley, Berlin-Heidelberg: Springer-Verlag; pp. 90-109), and the FLP/FRT system from the *Saccharomyces cerevisiae* 2  $\mu$  circle plasmid (Broach *et al.* *Cell* 29:227-234 (1982)).

Backman (U.S. Patent No. 4,673,640) discloses the *in vivo* use of  $\lambda$  recombinase to recombine a protein producing DNA segment by enzymatic site-specific recombination using wild-type recombination sites attB and attP.

Hasan and Szybalski (*Gene* 56:145-151 (1987)) discloses the use of  $\lambda$  Int recombinase *in vivo* for intramolecular recombination between wild type attP and attB sites which flank a promoter. Because the orientations of these sites are inverted relative to each other, this causes an irreversible flipping of the promoter region relative to the gene of interest.

Palazzolo *et al.* *Gene* 88:25-36 (1990), discloses phage lambda vectors having bacteriophage  $\lambda$  arms that contain restriction sites positioned outside a cloned DNA sequence and between wild-type loxP sites. Infection of *E. coli* cells that express the Cre recombinase with these phage vectors results in

recombination between the *loxP* sites and the *in vivo* excision of the plasmid replicon, including the cloned cDNA.

5 Pósfai *et al.* (*Nucl. Acids Res.* 22:2392-2398 (1994)) discloses a method for inserting into genomic DNA partial expression vectors having a selectable marker, flanked by two wild-type FRT recognition sequences. FLP site-specific recombinase as present in the cells is used to integrate the vectors into the genome at predetermined sites. Under conditions where the replicon is functional, this cloned genomic DNA can be amplified.

10 Bebee *et al.* (U.S. Patent No. 5,434,066) discloses the use of site-specific recombinases such as Cre for DNA containing two *loxP* sites is used for *in vivo* recombination between the sites.

35 Boyd (*Nucl. Acids Res.* 21:817-821 (1993)) discloses a method to facilitate the cloning of blunt-ended DNA using conditions that encourage intermolecular ligation to a dephosphorylated vector that contains a wild-type *loxP* site acted upon by a Cre site-specific recombinase present in *E. coli* host cells.

20 Waterhouse *et al.* (PCT No. 93/19172 and *Nucleic Acids Res.* 21 (9):2265 (1993)) disclose an *in vivo* method where light and heavy chains of a particular antibody were cloned in different phage vectors between *loxP* and *loxP 511* sites and used to transfect new *E. coli* cells. Cre, acting in the host cells on the two parental molecules (one plasmid, one phage), produced four products in equilibrium: two different cointegrates (produced by recombination at either *loxP* or *loxP 511* sites), and two daughter molecules, one of which was the desired product.

25 In contrast to the other related art, Schlake & Bode (*Biochemistry* 33:12746-12751 (1994)) discloses an *in vivo* method to exchange expression cassettes at defined chromosomal locations, each flanked by a wild type and a spacer-mutated FRT recombination site. A double-reciprocal crossover was mediated in cultured mammalian cells by using this FLP/FRT system for site-specific recombination.

**Transposases.** The family of enzymes, the transposases, has also been used to transfer genetic information between replicons. Transposons are structurally variable, being described as simple or compound, but typically encode the recombinase gene flanked by DNA sequences organized in inverted orientations. Integration of transposons can be random or highly specific. Representatives such as Tn7, which are highly site-specific, have been applied to the *in vivo* movement of DNA segments between replicons (Lucklow *et al.*, *J. Virol.* 67:4566-4579 (1993)).

Devine and Boeke *Nucl. Acids Res.* 22:3765-3772 (1994), discloses the construction of artificial transposons for the insertion of DNA segments, *in vitro*, into recipient DNA molecules. The system makes use of the integrase of yeast TY1 virus-like particles. The DNA segment of interest is cloned, using standard methods, between the ends of the transposon-like element TY1. In the presence of the TY1 integrase, the resulting element integrates randomly into a second target DNA molecule.

## DNA cloning

The cloning of DNA segments currently occurs as a daily routine in many research labs and as a prerequisite step in many genetic analyses. The purpose of these clonings is various, however, two general purposes can be considered: (1) the initial cloning of DNA from large DNA or RNA segments (chromosomes, YACs, PCR fragments, mRNA, etc.), done in a relative handful of known vectors such as pUC, pGem, pBlueScript, and (2) the subcloning of these DNA segments into specialized vectors for functional analysis. A great deal of time and effort is expended in the transfer of DNA segments from the initial cloning vectors to the more specialized vectors. This transfer is called subcloning.

The basic methods for cloning have been known for many years and have changed little during that time. A typical cloning protocol is as follows:

- (1) digest the DNA of interest with one or two restriction enzymes;
- (2) gel purify the DNA segment of interest when known;

(3) prepare the vector by cutting with appropriate restriction enzymes, treating with alkaline phosphatase, gel purify etc., as appropriate;

5 (4) ligate the DNA segment to the vector, with appropriate controls to eliminate background of uncut and self-ligated vector;

(5) introduce the resulting vector into an *E. coli* host cell;

(6) pick selected colonies and grow small cultures overnight;

(7) make DNA minipreps; and

10 (8) analyze the isolated plasmid on agarose gels (often after diagnostic restriction enzyme digestions) or by PCR.

The specialized vectors used for subcloning DNA segments are functionally diverse. These include but are not limited to: vectors for expressing genes in various organisms; for regulating gene expression; for providing tags to aid in protein purification or to allow tracking of proteins in cells; for modifying the cloned DNA segment (e.g., generating deletions); for the synthesis of probes (e.g., riboprobes); for the preparation of templates for DNA sequencing; for the identification of protein coding regions; for the fusion of various protein-coding regions; to provide large amounts of the DNA of interest, etc. It is common that a particular investigation will involve subcloning the DNA segment of interest into several different specialized vectors.

25 As known in the art, simple subclonings can be done in one day (e.g., the DNA segment is not large and the restriction sites are compatible with those of the subcloning vector). However, many other subclonings can take several weeks, especially those involving unknown sequences, long fragments, toxic genes, unsuitable placement of restriction sites, high backgrounds, impure enzymes, etc. Subcloning DNA fragments is thus often viewed as a chore to be done as few times as possible. Several methods for facilitating the cloning of DNA segments have been described, e.g., as in the following references.

30 Ferguson, J., et al. *Gene* 16:191 (1981), discloses a family of vectors for subcloning fragments of yeast DNA. The vectors encode kanamycin resistance. Clones of longer yeast DNA segments can be partially digested and ligated into

the subcloning vectors. If the original cloning vector conveys resistance to ampicillin, no purification is necessary prior to transformation, since the selection will be for kanamycin.

5 Hashimoto-Gotoh, T., *et al. Gene* 41:125 (1986), discloses a subcloning vector with unique cloning sites within a streptomycin sensitivity gene; in a streptomycin-resistant host, only plasmids with inserts or deletions in the dominant sensitivity gene will survive streptomycin selection.

10 Accordingly, traditional subcloning methods, using restriction enzymes and ligase, are time consuming and relatively unreliable. Considerable labor is expended, and if two or more days later the desired subclone can not be found among the candidate plasmids, the entire process must then be repeated with alternative conditions attempted. Although site specific recombinases have been used to recombine DNA *in vivo*, the successful use of such enzymes *in vitro* was expected to suffer from several problems. For example, the site specificities and efficiencies were expected to differ *in vitro*; topologically-linked products were expected; and the topology of the DNA substrates and recombination proteins was expected to differ significantly *in vitro* (see, e.g., Adams *et al. J. Mol. Biol.* 226:661-73 (1992)). Reactions that could go on for many hours *in vivo* were expected to occur in significantly less time *in vitro* before the enzymes became inactive. Multiple DNA recombination products were expected in the biological host used, resulting in unsatisfactory reliability, specificity or efficiency of subcloning. Thus, *in vitro* recombination reactions were not expected to be sufficiently efficient to yield the desired levels of product.

25 **Ribosomal Proteins**

**Characterization.** *E. coli* ribosomes have some 53 different proteins, 21 associated with the 30S subunit (designated S1 through S21) and 32 associated with the 50S subunit (designated L1 through L34). Generally, the lower the number the higher the molecular weight. With the exception of S1 through S4 and L1 through L4, they contain less than 200 amino acids (molecular weights are less than 20 KDa). The primary amino acid sequence of each protein is known.

The three-dimensional structures of S5, S6, S8, S17, L1, L7, L9, L14, and L30 are known. Most of these proteins have a relatively high proportion of the two basic amino acids arginine (arg or R) and lysine (lys or K). This intuitively makes sense if most of the ribosomal proteins are assumed to be RNA binding proteins.

5 Much of what is known about ribosomal proteins has been summarized in a series of articles in *Annual Reviews of Biochemistry*: 51:155 (1982); 52:35 (1983); 53:75 (1984); 54:507 (1985); 66:679 (1997).

10 **Enhancement of Yeast Recombination Systems.** The yeast FLP/FRT recombination system requires only the FRT DNA binding site and FLP recombinase to carry out recombination. In contrast, the minimum requirements for carrying out recombination in the  $\lambda$  integrase (Int) system include a recombinase (Int) and DNA sites (*att*), but also IHF protein. IHF is a member of the HU family of small DNA binding proteins. These are basic proteins of 100 amino acids or less that bind to DNA and condense its structure. HU will substitute for IHF in the  $\lambda$  recombination system. While IHF and HU do not stimulate the yeast FLP/FRT recombination system, the *E. coli* ribosomal proteins S3, S4, S5, and L2 do (Bruckner and Cox, *Nucl. Acids Res.* 17:3145-3161 (1989)). The *E. coli* ribosomal proteins that have been shown to stimulate the yeast FLP/FRT recombination system are large, all possessing, with one exception, more than 200 amino acids (Table 1); smaller *E. coli* ribosomal proteins have not been shown to stimulate the FLP/FRT (or any other) recombination system.

15  
20

TABLE I

*E. coli* RIBOSOMAL PROTEINS THAT STIMULATE  
YEAST FLP/FRT RECOMBINASE

5

10

0  
5  
10  
15  
20  
25  
30

<i>E. coli</i> Ribosomal Protein	No. of Basic Residues (Percentage of Total)	Total No. of Residues	Molec. Weight
S3	39 (16.8%)	232	25,852
S4	39 (19.2%)	203	23,137
S5	22 (13.3%)	166	17,515
L2	48 (17.8%)	269	29,416

## SUMMARY OF THE INVENTION

The present invention provides compositions and methods for obtaining amplified, chimeric or recombinant nucleic acid molecules using recombinational cloning, *in vitro* or *in vivo*. These methods are highly specific, rapid, and less labor intensive than standard cloning or subcloning techniques. The improved specificity, speed and yields of the present invention facilitates DNA or RNA cloning or subcloning, regulation or exchange useful for any related purpose.

25

30

In one embodiment, the present invention relates to compositions for use in cloning or subcloning one or more desired nucleic acid molecules by recombinational cloning, comprising at least one ribosomal protein and at least one recombination protein. In a related aspect, the compositions may comprise more than one ribosomal protein and/or more than one recombination protein. Preferably, prokaryotic ribosomal proteins and prokaryotic recombination proteins are used, although eukaryotic ribosomal proteins and/or eukaryotic recombination proteins may also function in accordance with the invention. According to the invention, the ribosomal proteins used may be basic ribosomal proteins, and may be no larger than about 14 kilodaltons in size.

In certain preferred embodiments, the ribosomal protein may be a prokaryotic ribosomal protein, such as an *Escherichia coli* ribosomal protein, particularly an *E. coli* protein including but not limited to S10, S14, S15, S16, S17, S18, S19, S20, S21, L21, L23, L24, L25, L27, L28, L29, L30, L31, L32, L33 and L34, and most particularly S20, L27 and/or S15. In related embodiments, the recombination protein for use in the compositions is selected from the group consisting of Int, Cre, FLP, Xis, IHF and HU, and is preferably Int. These compositions of the invention may further comprise one or more nucleic acid molecules, including but not limited to one or more Insert Donor molecules, one or more Vector Donor molecules, one or more cointegrate molecules, one or more Product molecules and one or more Byproduct molecules.

The invention also relates generally to methods of cloning or subcloning one or more desired nucleic acid molecules by recombinational cloning. In one such aspect, the invention relates to such methods comprising:

- (a) combining *in vitro* or *in vivo*
  - (i) one or more Insert Donor molecules comprising one or more desired nucleic acid segments flanked by at least two recombination sites, wherein the recombination sites do not substantially recombine with each other;
  - (ii) one or more Vector Donor molecules comprising at least two recombination sites, wherein the recombination sites do not substantially recombine with each other;
  - (iii) at least one recombination protein; and
  - (iv) at least one ribosomal protein;
- (b) incubating the combination formed in step (a) under conditions sufficient to transfer one or more of the desired segments into one or more of the Vector Donor molecules, thereby producing one or more desired Product nucleic acid molecules;

and optionally:

- (c) combining *in vitro* or *in vivo*

5

- (i) one or more of the Product molecules comprising the desired segments flanked by two or more recombination sites, wherein the recombination sites do not substantially recombine with each other;
- (ii) one or more different Vector Donor molecules comprising two or more recombination sites, wherein the recombination sites do not substantially recombine with each other;
- (iii) at least one recombination protein; and
- (iv) at least one ribosomal protein; and

10

- (d) incubating the combination formed in step (c) under conditions sufficient to transfer one or more of the desired segments into one or more different Vector Donor molecules, thereby producing one or more different Product molecules.

The invention also relates to such methods which further comprise incubating the different Product molecules with one or more different Vector Donor molecules under conditions sufficient to transfer one or more of the desired segments into the different Vector Donor molecules.

In a related aspect, the invention relates to methods of cloning or subcloning one or more desired nucleic acid molecules by recombinational cloning comprising:

20

- a) combining *in vitro* or *in vivo*
  - i) one or more Insert Donor molecules comprising one or more nucleic acid segments flanked by two or more recombination sites, wherein the recombination sites do not substantially recombine with each other;
  - 25 ii) two or more different Vector Donor molecules comprising two or more recombination sites, wherein the recombination sites do not substantially recombine with each other;
  - 30 iii) at least one recombination protein; and

5

- iv) at least one ribosomal protein; and
- b) incubating the combination formed in step (a) under conditions sufficient to transfer one or more of the desired segments into the different Vector Donor molecules, thereby producing two or more different Product molecules.

10 According to the invention, the one or more ribosomal proteins and the one or more recombination proteins for use in these methods are preferably those prokaryotic and/or eukaryotic ribosomal and recombination proteins described herein for use in the compositions of the invention.

15 In another related aspect, the invention relates to methods of cloning or subcloning one or more desired nucleic acid molecules by recombinational cloning comprising:

20

- (a) combining *in vitro* or *in vivo*
  - (i) one or more Insert Donor molecules comprising one or more desired nucleic acid segments flanked by at least two recombination sites, wherein the recombination sites do not substantially recombine with each other;
  - (ii) one or more Vector Donor molecules comprising at least two recombination sites, wherein the recombination sites do not substantially recombine with each other; and
  - (iii) one or more of the compositions of the invention;
- (b) incubating the combination formed in step (a) under conditions sufficient to transfer one or more of the desired segments into one or more of the Vector Donor molecules, thereby producing one or more desired Product nucleic acid molecules;

25 and optionally:

30

- (c) combining *in vitro* or *in vivo*
  - (i) one or more of the Product molecules comprising the desired segments flanked by two or more recombination sites, wherein the recombination sites do not substantially recombine with each other;

- (ii) one or more different Vector Donor molecules comprising two or more recombination sites, wherein the recombination sites do not substantially recombine with each other; and
- (iii) one or more of the compositions of the invention; and

(d) incubating the combination formed in step (c) under conditions sufficient to transfer one or more of the desired segments into one or more different Vector Donor molecules, thereby producing one or more different Product molecules.

10 In another related aspect, the invention relates to methods of cloning or  
subcloning one or more desired nucleic acid molecules by recombinational  
cloning comprising:

- a) combining *in vitro* or *in vivo*
  - i) one or more Insert Donor molecules comprising one or more nucleic acid segments flanked by two or more recombination sites, wherein the recombination sites do not substantially recombine with each other;
  - ii) two or more different Vector Donor molecules comprising two or more recombination sites, wherein the recombination sites do not substantially recombine with each other; and
  - iii) one or more of the compositions of the invention; and
- b) incubating the combination formed in step (a) under conditions sufficient to transfer one or more of the desired segments into the different Vector Donor molecules, thereby producing two or more different Product molecules.

In another related aspect, the invention relates to methods for recombinational cloning of one or more desired nucleic acid molecules comprising

(a) mixing one or more desired nucleic acid molecules with one or more vectors and with one or more of the compositions of the invention; and

(b) incubating the mixture under conditions sufficient to transfer the one or more desired nucleic acid molecules into one or more of the vectors.

In another related aspect, the invention relates to methods for enhancement of recombinational cloning of nucleic acid molecules, comprising contacting one or more nucleic acid molecules with one or more ribosomal proteins and one or more recombination proteins, or with one or more compositions of the invention, under conditions favoring the recombinational cloning of the one or more nucleic acid molecules.

According to the invention, the Insert Donor molecules and nucleic acid molecules for use in the compositions and methods of the invention may be derived from genomic DNA or cDNA, or may be produced by chemical synthesis methods. In a related aspect, the Insert Donor molecules may comprise one or more vectors.

According to the invention, the Vector Donor molecules for use in the compositions and methods of the invention may comprise at least one Selectable marker, which may be an antibiotic resistance gene, a tRNA gene, an auxotrophic marker, a toxic gene, a phenotypic marker, an antisense oligonucleotide, a restriction endonuclease, a restriction endonuclease cleavage site, an enzyme cleavage site, a protein binding site, and a sequence complementary to a PCR primer sequence. In a related aspect, the Vector Donor molecules may comprise one or more eukaryotic vectors or one or more prokaryotic vectors. Eukaryotic vectors suitable for use in this aspect of the invention may comprise, for example, vectors which propagate and/or replicate in yeast cells, plant cells, fish cells, eukaryotic cells, mammalian cells, and/or insect cells, while suitable prokaryotic vectors may comprise, for example, vectors which propagate and/or replicate in bacteria of the genera *Escherichia* (most particularly *E. coli*), *Salmonella*, *Bacillus*, *Serratia*, *Streptomyces* or *Pseudomonas*.

The invention also relates generally to DNA molecules produced by the methods of the invention, particularly to such DNA molecules which are isolated DNA molecules. The invention also relates to vectors comprising such DNA molecules, and to host cells comprising such DNA molecules and/or vectors.

The invention also relates to kits for use in recombinational cloning of a nucleic acid molecule. In one such aspect, the kits of the invention may comprise one or more containers, particularly wherein the kit contains at least one ribosomal protein and at least one recombination protein. Such proteins may be contained in separate containers in the kit, or may be combined into a common container or containers. In a related aspect, the kits of the invention may comprise combinations of different ribosomal proteins and/or combinations of different recombination proteins. Ribosomal proteins and recombination proteins suitable for use in the kits of the invention include, but are not necessarily limited to, those described in detail herein.

Other preferred embodiments of the present invention will be apparent to one of ordinary skill in light of what is known in the art, the following drawings and description of the invention, and the claims.

#### BRIEF DESCRIPTION OF THE DRAWINGS

**Figure 1** depicts one general method of the present invention, wherein the starting (parent) DNA molecules can be circular or linear. The goal is to exchange the new subcloning vector D for the original cloning vector B. It is desirable in one embodiment to select for AD and against all the other molecules, including the Cointegrate. The square and circle are sites of recombination: *e.g.*, *loxP* sites, *att* sites, *etc.* For example, segment D can contain expression signals, new drug markers, new origins of replication, or specialized functions for mapping or sequencing DNA.

**Figure 2** depicts a restriction map for plasmid pHN894. *AttP*: *attP* attachment site; *tet*: truncated tetracycline resistance gene; *amp*:  $\beta$ -lactamase gene.

**Figure 3** depicts a restriction map for plasmid pBB105. *attB*: *attB* attachment site; *tet*: truncated tetracycline resistance gene; *amp*:  $\beta$ -lactamase gene; *ori*: *colE1* origin of replication; *ROP*: replication control site.

5 **Figure 4** depicts a restriction map for plasmid pHN872. attL: *attL* attachment site; 'tet: truncated tetracycline resistance gene; 'amp: truncated β-lactamase gene; ori: colE1 origin of replication; KmR: kanamycin resistance gene.

10 **Figure 5** depicts a restriction map for plasmid pHN868. attR: *attR* attachment site; 'tet: truncated tetracycline resistance gene; amp: β-lactamase gene; ori: colE1 origin of replication; ROP: replication control site.

15 **Figure 6** depicts a restriction map for plasmid pEZ13835. WTattP1: modified *attP* attachment site; WTattP3: modified *attP* attachment site; T1T2: transcription terminators; KmR: kanamycin resistance gene; CmR: chloramphenicol resistance gene; ccdB: death gene; ori: colE1 origin of replication.

20 **Figure 7** depicts a restriction map for plasmid pEYC7501. attB1: modified *attB* attachment site; attB3: modified *attB* attachment site; GFP: truncated green fluorescent protein gene; T7 P: T7 promoter; SP6 P: SP6 promoter; CMV P: CMV promoter; lacI': lacI promoter; lox p: cre recombination site; small t & poly A: SV40 small tumor antigen intron and poly A signal; f1: f1 intergenic region; incA: phage P1 incompatibility locus; Amp: β-lactamase gene; ori: colE1 origin of replication.

25 **Figure 8** depicts a restriction map for plasmid pEZ11104. attL1: modified *attL* attachment site; attL3: modified *attL* attachment site; CmR: chloramphenicol resistance gene; KmR: kanamycin resistance gene; ori: colE1 origin of replication.

30 **Figure 9** depicts a restriction map for plasmid pEYC8402. attR1: modified *attR* attachment site; attR3: modified *attR* attachment site; lac I: lac repressor gene; amp: β-lactamase gene; ori: colE1 origin of replication; CmR: chloramphenicol resistance gene; f1: f1 intergenic region; ccdB: death gene.

**Figure 10** depicts a restriction map for plasmid pTRCN2. Ap: β-lactamase gene; ptrc: trc promoter; lacI<sup>Q</sup>: lac repressor gene; f1'ori: f1 intergenic region; ori: colE1 origin of replication.

5 **Figure 11** depicts a restriction map for plasmid pTRCN2INT2. Ap:  $\beta$ -lactamase gene; ptrc: trc promoter; lacI<sup>Q</sup>: lac repressor gene; fl'ori: fl intergenic region; ori: colE1 origin of replication; Int:  $\lambda$  integrase gene.

10 **Figure 12** depicts a restriction map for plasmid pTRCN2XIS1. Ap:  $\beta$ -lactamase gene; ptrc: trc promoter; lacI<sup>Q</sup>: lac repressor gene; fl'ori: fl intergenic region; ori: colE1 origin of replication; xis:  $\lambda$  xis gene.

15 **Figure 13** depicts a restriction map for plasmid pTRCN2S20AA. Ap:  $\beta$ -lactamase gene; ptrc: trc promoter; lacI<sup>Q</sup>: lac repressor gene; fl'ori: fl intergenic region; ori: colE1 origin of replication; rpsT: S20 gene.

20 **Figure 14** depicts a restriction map for plasmid pET12AS20AA. Ap:  $\beta$ -lactamase gene; ori: colE1 origin of replication; rpsT: S20 gene; T7: T7 promoter; T7 term: T7 transcription termination sequence.

25 **Figure 15** is a photograph of an SDS-PAGE gel of fractions from phosphocellulose column fractionation of proteins not bound by hydroxyapatite. Aliquots (7.5  $\mu$ l) from fractions 13 through 20 of the phosphocellulose column of proteins not bound by hydroxyapatite were analyzed by SDS PAGE. IHF ("IHF A": 0.3  $\mu$ g; "IHF B": 0.5  $\mu$ g) and BenchMark protein standards ("M") were run as references. The bottom of the figure indicates the relative ability of aliquots from the fractions to stimulate Int in an integrative recombination gel assay (-, no stimulation; +, ++, +++, increasing levels of stimulation).

30 **Figure 16** is a photograph of an SDS-PAGE gel of S20 ribosomal protein purified from a side fraction of a native Int purification. Lanes M: BenchMark protein standards; lanes A through E: 5-, 2-, 2-, 1-, and 1- $\mu$ l aliquots, respectively, of Mono S pool of S20.

35 **Figure 17** is a photograph of an ethidium bromide-stained gel in an integrative recombination gel assay (see Materials and Methods) showing the ability of S20 protein in the Mono S pool (see Figure 16) to stimulate Int activity. Lane A: Int plus S20; lane B: Int alone; lane C: Int dilution buffer alone. The slowest migrating band is the recombinant DNA product.

40 **Figure 18** is a photograph of an SDS-PAGE gel of peak fractions containing integrative recombination stimulatory activity from the Mono S

columns described in Materials and Methods section Purification of Stimulatory Proteins from Cells producing Native Int and Results section PART II: Purification and Identification of the Stimulatory Proteins. Phosphocellulose Pool #1 was fractionated on a Mono S column producing two peaks of activity at fraction 18 (1 and 2  $\mu$ l, lanes A and B) and fraction 22 (1 and 2  $\mu$ l, lanes C and D). Phosphocellulose Pool #2 was fractionated in a second run on the same Mono S column producing one peak of activity at fraction 24 (1 and 2  $\mu$ l, lanes F and G). S20 was run in lane E and BenchMark protein standard in lane M.

**Figure 19** is a photograph of an ethidium bromide-stained gel in an integrative recombination gel assay (Materials and Methods) showing stimulation of 37 ng of native Int by 900 ng of recombinant S20 (Figure 19), 900 ng of S20 (see Figure 16), and 10  $\mu$ g of L27 (fraction 18 in Figure 18). Lane A: recombinant S20; lane B: S20; lane C: L27; lane D: Int alone; lane E: no added Int or stimulatory protein.

**Figure 20** is a photograph of an SDS-PAGE gel of 2  $\mu$ g of purified recombinant S20.

**Figure 21** is a photograph of an ethidium bromide-stained gel in integrative (lanes A to C) and excisive (lanes D to F) recombination gel assays, showing the recombinase activity of 59 ng of Int-His<sub>6</sub> in the presence of 0 ng (lanes B and E) and 382 ng (lanes C and F) of recombinant S20. All assays also contained 12.5 ng IHF. Excisive recombination assays contained 42 ng Xis-His<sub>6</sub>. The assays analyzed in lanes A and D contained no Int-His<sub>6</sub> or rS20.

DETAILED DESCRIPTION OF THE INVENTION

**Overview**

It has been unexpectedly discovered by the present invention that one or more ribosomal proteins, which may be one or more prokaryotic or eukaryotic ribosomal proteins and particularly one or more *E. coli* ribosomal proteins, may be used to enhance, stimulate, or restore the *in vitro* and *in vivo* recombination activity of recombination systems, which may be prokaryotic or eukaryotic recombination systems, such as the  $\lambda$  Int recombination system. Thus, the invention provides compositions comprising such ribosomal proteins, and methods using such compositions, which are useful in performing reversible and/or repeatable cloning and subcloning reactions to manipulate nucleic acid molecules in order to form chimeric nucleic acids using recombination proteins (e.g.,  $\lambda$  Int) and recombination sites. Recombinational cloning according to the present invention thus uses compositions comprising one or more ribosomal proteins, and one or more recombination proteins (which may be site-specific prokaryotic recombination proteins), in combination with recombinant nucleic acid molecules having at least one selected recombination site for moving or exchanging segments of nucleic acid molecules, *in vitro* and *in vivo*.

The methods of the invention use recombination reactions to generate chimeric DNA or RNA molecules that have the desired characteristic(s) and/or nucleic acid segment(s). The methods of the invention function such that a nucleic acid molecule of interest may be moved or transferred into any number of vector systems. In accordance with the invention, such transfer to various vector systems may be accomplished separately, sequentially or in mass (e.g. into any number of different vectors in one step). The improved specificity, speed and/or yields of the present invention facilitates DNA or RNA cloning, subcloning, regulation or exchange useful for any related purpose. Such purposes include *in vitro* recombination of DNA or RNA segments and *in vitro* or *in vivo* insertion or modification of transcribed, replicated, isolated or genomic DNA or RNA.

5                   **Definitions**

In the description that follows, a number of terms used in recombinant DNA technology are utilized extensively. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided.

10                   **Adapter:** is an oligonucleotide or nucleic acid fragment or segment (preferably DNA) which comprises one or more recombination sites (or portions of such recombination sites) which in accordance with the invention can be added to a circular or linear Insert Donor molecule as well as other nucleic acid molecules described herein. When using portions of recombination sites, the missing portion may be provided by the Insert Donor molecule. Such adapters may be added at any location within a circular or linear molecule, although the adapters are preferably added at or near one or both termini of a linear molecule. Preferably, adapters are positioned to be located on both sides (flanking) a particularly nucleic acid molecule of interest. In accordance with the invention, adapters may be added to nucleic acid molecules of interest by standard recombinant techniques (e.g. restriction digest and ligation). For example, adapters may be added to a circular molecule by first digesting the molecule with an appropriate restriction enzyme, adding the adapter at the cleavage site and reforming the circular molecule which contains the adapter(s) at the site of cleavage. Alternatively, adapters may be ligated directly to one or more and preferably both termini of a linear molecule thereby resulting in linear molecule(s) having adapters at one or both termini. In one aspect of the invention, adapters may be added to a population of linear molecules, (e.g. a cDNA library or genomic DNA which has been cleaved or digested) to form a population of linear molecules containing adapters at one and preferably both termini of all or substantial portion of said population.

15                   **Amplification:** refers to any *in vitro* method for increasing a number of copies of a nucleotide sequence with the use of a polymerase. Nucleic acid amplification results in the incorporation of nucleotides into a DNA and/or RNA molecule or primer thereby forming a new molecule complementary to a

template. The formed nucleic acid molecule and its template can be used as templates to synthesize additional nucleic acid molecules. As used herein, one amplification reaction may consist of many rounds of replication. DNA amplification reactions include, for example, polymerase chain reaction (PCR). One PCR reaction may consist of 5-100 "cycles" of denaturation and synthesis of a DNA molecule.

**Byproduct:** is a daughter molecule (a new clone produced after the second recombination event during the recombinational cloning process) lacking the segment which is desired to be cloned or subcloned.

**Cointegrate:** is at least one recombination intermediate nucleic acid molecule of the present invention that contains both parental (starting) molecules. It will usually be circular. In some embodiments it can be linear.

**Host:** is any prokaryotic or eukaryotic organism that can be a recipient of the recombinational cloning Product. A "host," as the term is used herein, includes prokaryotic or eukaryotic organisms that can be genetically engineered. For examples of such hosts, see Maniatis *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1982).

**Hybridization:** The terms "hybridization" and "hybridizing" refers to base pairing of two complementary single-stranded nucleic acid molecules (RNA and/or DNA) to give a double stranded molecule. As used herein, two nucleic acid molecules may be hybridized, although the base pairing is not completely complementary. Accordingly, mismatched bases do not prevent hybridization of two nucleic acid molecules provided that appropriate conditions, well known in the art, are used.

**Insert or Inserts:** include the desired nucleic acid segment or a population of nucleic acid segments (segment *A* of Figure 1) which may be manipulated by the methods of the present invention. Thus, the terms Insert(s) are meant to include a particular nucleic acid (preferably DNA) segment or a population of segments. Such Insert(s) can comprise one or more genes.

**Insert Donor:** is one of the two parental nucleic acid molecules (e.g. RNA or DNA) of the present invention which carries the Insert. The Insert Donor

5 molecule comprises the Insert flanked on both sides with recombination sites. The Insert Donor can be linear or circular. In one embodiment of the invention, the Insert Donor is a circular DNA molecule and further comprises a cloning vector sequence outside of the recombination signals (see Figure 1). When a population of Inserts or population of nucleic acid segments are used to make the Insert Donor, a population of Insert Donors result and may be used in accordance with the invention.

10 **Library:** refers to a collection of nucleic acid molecules (circular or linear). In one preferred embodiment, a library is representative of all or a significant portion of the DNA content of an organism (a "genomic" library), or a set of nucleic acid molecules representative of all or a significant portion of the expressed genes (a cDNA library) in a cell, tissue, organ or organism. A library may also comprise random sequences made by *de novo* synthesis, mutagenesis of one or more sequences and the like. Such libraries may or may not be contained in one or more vectors.

15 **Nucleotide:** refers to a base-sugar-phosphate combination. Nucleotides are monomeric units of a nucleic acid sequence (DNA and RNA). The term nucleotide includes ribonucleoside triphosphatase ATP, UTP, CTG, GTP and deoxyribonucleoside triphosphates such as dATP, dCTP, dITP, dUTP, dGTP, dTTP, or derivatives thereof. Such derivatives include, for example, [ $\alpha$ S]dATP, 7-deaza-dGTP and 7-deaza-dATP. The term nucleotide as used herein also refers to dideoxyribonucleoside triphosphates (ddNTPs) and their derivatives. Illustrated examples of dideoxyribonucleoside triphosphates include, but are not limited to, ddATP, ddCTP, ddGTP, ddITP, and ddTTP. According to the present invention, a "nucleotide" may be unlabeled or detectably labeled by well known techniques. Detectable labels include, for example, radioactive isotopes, fluorescent labels, chemiluminescent labels, bioluminescent labels and enzyme labels.

25 **Oligonucleotide:** refers to a synthetic or natural molecule comprising a covalently linked sequence of nucleotides which are joined by a phosphodiester

bond between the 3' position of the deoxyribose or ribose of one nucleotide and the 5' position of the deoxyribose or ribose of the adjacent nucleotide.

**Primer:** refers to a single stranded or double stranded oligonucleotide that is extended by covalent bonding of nucleotide monomers during amplification or polymerization of a nucleic acid molecule (e.g. a DNA molecule). In a preferred aspect, the primer comprises one or more recombination sites or portions of such recombination sites. Portions of recombination sites comprise at least 2 bases, at least 5 bases, at least 10 bases or at least 20 bases of the recombination sites of interest. When using portions of recombination sites, the missing portion of the recombination site may be provided by the newly synthesized nucleic acid molecule. Such recombination sites may be located within and/or at one or both termini of the primer. Preferably, additional sequences are added to the primer adjacent to the recombination site(s) to enhance or improve recombination and/or to stabilize the recombination site during recombination. Such stabilization sequences may be any sequences (preferably G/C rich sequences) of any length. Preferably, such sequences range in size from 1 to about 1000 bases, 1 to about 500 bases, and 1 to about 100 bases, 1 to about 60 bases, 1 to about 25, 1 to about 10, 2 to about 10 and preferably about 4 bases. Preferably, such sequences are greater than 1 base in length and preferably greater than 2 bases in length.

**Product:** is one the desired daughter molecules comprising the **A** and **D** sequences which is produced after the second recombination event during the recombinational cloning process (see Figure 1). The Product contains the nucleic acid which was to be cloned or subcloned. In accordance with the invention, when a population of Insert Donors are used, the resulting population of Product molecules will contain all or a portion of the population of Inserts of the Insert Donors and preferably will contain a representative population of the original molecules of the Insert Donors.

**Promoter:** is a DNA sequence generally described as the 5'-region of a gene, located proximal to the start codon. The transcription of an adjacent DNA segment is initiated at the promoter region. A repressible promoter's rate of

transcription decreases in response to a repressing agent. An inducible promoter's rate of transcription increases in response to an inducing agent. A constitutive promoter's rate of transcription is not specifically regulated, though it can vary under the influence of general metabolic conditions.

**Recognition sequence:** Recognition sequences are particular sequences which a protein, chemical compound, DNA, or RNA molecule (e.g., restriction endonuclease, a modification methylase, or a recombinase) recognizes and binds. In the present invention, a recognition sequence will usually refer to a recombination site. For example, the recognition sequence for Cre recombinase is *loxP* which is a 34 base pair sequence comprised of two 13 base pair inverted repeats (serving as the recombinase binding sites) flanking an 8 base pair core sequence. See Figure 1 of Sauer, B., *Current Opinion in Biotechnology* 5:521-527 (1994). Other examples of recognition sequences are the *attB*, *attP*, *attL*, and *attR* sequences which are recognized by the recombinase enzyme  $\lambda$  Integrase. *attB* is an approximately 25 base pair sequence containing two 9 base pair core-type Int binding sites and a 7 base pair overlap region. *attP* is an approximately 240 base pair sequence containing core-type Int binding sites and arm-type Int binding sites as well as sites for auxiliary proteins integration host factor (IHF), FIS, and excisionase (Xis). See Landy, *Current Opinion in Biotechnology* 3:699-707 (1993). Such sites may also be engineered according to the present invention to enhance production of products in the methods of the invention. When such engineered sites lack the P1 or H1 domains to make the recombination reactions irreversible (e.g., *attR* or *attP*), such sites may be designated *attR'* or *attP'* to show that the domains of these sites have been modified in some way.

**Recombinase:** is a type of recombination protein which catalyzes the exchange of DNA segments at specific recombination sites.

**Recombinational Cloning:** is a method described herein, whereby segments of nucleic acid molecules or populations of such molecules are exchanged, inserted, replaced, substituted or modified, *in vitro* or *in vivo*.

**Recombination proteins:** include excisive or integrative proteins, enzymes, co-factors or associated proteins that are involved in recombination reactions involving one or more recombination sites. *See, Landy (1994), infra.*

**5 Repression cassette:** is a nucleic acid segment that contains a repressor of a Selectable marker present in the subcloning vector.

**10 Ribosomal protein:** is a polypeptide, protein, or a functional fragment, mutant, or derivative thereof, that is a constituent of a subunit of a ribosome. According to the invention, the ribosome may be a prokaryotic or eukaryotic ribosome, and is preferably a prokaryotic ribosome, particularly an *E. coli* ribosome, comprising a 30S and a 50S subunit. By a "functional" fragment, mutant, or derivative thereof is meant a fragment, mutant, or derivative of a native ribosomal protein that has substantially the same biological activity as the corresponding native ribosomal protein in stimulating a recombination system such as the  $\lambda$  Int recombination system.

**20 Selectable marker:** is a DNA segment that allows one to select for or against a molecule or a cell that contains it, often under particular conditions. These markers can encode an activity, such as, but not limited to, production of RNA, peptide, or protein, or can provide a binding site for RNA, peptides, proteins, inorganic and organic compounds or compositions and the like. Examples of Selectable markers include but are not limited to: (1) DNA segments that encode products which provide resistance against otherwise toxic compounds (e.g., antibiotics); (2) DNA segments that encode products which are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); (3) DNA segments that encode products which suppress the activity of a gene product; (4) DNA segments that encode products which can be readily identified (e.g., phenotypic markers such as  $\beta$ -galactosidase, green fluorescent protein (GFP), and cell surface proteins); (5) DNA segments that bind products which are otherwise detrimental to cell survival and/or function; (6) DNA segments that otherwise inhibit the activity of any of the DNA segments described in Nos. 1-5 above (e.g., antisense oligonucleotides); (7) DNA segments that bind products that modify a substrate (e.g. restriction endonucleases); (8) DNA segments that

can be used to isolate or identify a desired molecule (e.g. specific protein binding sites); (9) DNA segments that encode a specific nucleotide sequence which can be otherwise non-functional (e.g., for PCR amplification of subpopulations of molecules); (10) DNA segments, which when absent, directly or indirectly confer resistance or sensitivity to particular compounds; and/or (11) DNA segments that encode products which are toxic in recipient cells.

**Selection scheme:** is any method which allows selection, enrichment, or identification of a desired Product or Product(s) from a mixture containing the Insert Donor, Vector Donor, any intermediates (e.g. a Cointegrate), and/or Byproducts. The selection schemes of one preferred embodiment have at least two components that are either linked or unlinked during recombinational cloning. One component is a Selectable marker. The other component controls the expression *in vitro* or *in vivo* of the Selectable marker, or survival of the cell harboring the plasmid carrying the Selectable marker. Generally, this controlling element will be a repressor or inducer of the Selectable marker, but other means for controlling expression of the Selectable marker can be used. Whether a repressor or activator is used will depend on whether the marker is for a positive or negative selection, and the exact arrangement of the various DNA segments, as will be readily apparent to those skilled in the art. A preferred requirement is that the selection scheme results in selection of or enrichment for only one or more desired Products. As defined herein, selecting for a DNA molecule includes (a) selecting or enriching for the presence of the desired DNA molecule, and (b) selecting or enriching against the presence of DNA molecules that are not the desired DNA molecule.

In one embodiment, the selection schemes (which can be carried out in reverse) will take one of three forms, which will be discussed in terms of Figure 1. The first, exemplified herein with a Selectable marker and a repressor therefore, selects for molecules having segment **D** and lacking segment **C**. The second selects against molecules having segment **C** and for molecules having segment **D**. Possible embodiments of the second form would have a DNA segment carrying a gene toxic to cells into which the *in vitro* reaction products are

to be introduced. A toxic gene can be a DNA that is expressed as a toxic gene product (a toxic protein or RNA), or can be toxic in and of itself. (In the latter case, the toxic gene is understood to carry its classical definition of "heritable trait".)

5 Examples of such toxic gene products are well known in the art, and include, but are not limited to, restriction endonucleases (e.g., *Dpn*I), apoptosis-related genes (e.g. ASK1 or members of the bcl-2/ced-9 family), retroviral genes including those of the human immunodeficiency virus (HIV), defensins such as NP-1, inverted repeats or paired palindromic DNA sequences, bacteriophage lytic genes such as those from  $\phi$ X174 or bacteriophage T4; antibiotic sensitivity genes such as *rps*L, antimicrobial sensitivity genes such as *phe*S, plasmid killer genes, eukaryotic transcriptional vector genes that produce a gene product toxic to bacteria, such as GATA-1, and genes that kill hosts in the absence of a suppressing function, e.g., *kicB* or *ccdB*. A toxic gene can alternatively be selectable *in vitro*, e.g., a restriction site.

10 15 20 Many genes coding for restriction endonucleases operably linked to inducible promoters are known, and may be used in the present invention. See, e.g. U.S. Patent Nos. 4,960,707 (*Dpn*I and *Dpn*II); 5,000,333, 5,082,784 and 5,192,675 (*Kpn*I); 5,147,800 (*Ngo*AI<sub>III</sub> and *Ngo*AI); 5,179,015 (*Fsp*I and *Hae*III); 5,200,333 (*Hae*II and *Tag*I); 5,248,605 (*Hpa*II); 5,312,746 (*Cl*I); 5,231,021 and 5,304,480 (*Xho*I and *Xho*II); 5,334,526 (*Alu*I); 5,470,740 (*Nsi*I); 5,534,428 (*Sst*I/*Sac*I); 5,202,248 (*Nco*I); 5,139,942 (*Nde*I); and 5,098,839 (*Pac*I). See also Wilson, G.G., *Nucl. Acids Res.* 19:2539-2566 (1991); and Lunnen, K.D., *et al.*, *Gene* 74:25-32 (1988).

25 In the second form, segment **D** carries a Selectable marker. The toxic gene would eliminate transformants harboring the Vector Donor, Cointegrate, and Byproduct molecules, while the Selectable marker can be used to select for cells containing the Product and against cells harboring only the Insert Donor.

30 The third form selects for cells that have both segments **A** and **D** in *cis* on the same molecule, but not for cells that have both segments in *trans* on different

5 molecules. This could be embodied by a Selectable marker that is split into two inactive fragments, one each on segments **A** and **D**.

The fragments are so arranged relative to the recombination sites that when the segments are brought together by the recombination event, they reconstitute a functional Selectable marker. For example, the recombinational event can link a promoter with a structural gene, can link two fragments of a structural gene, or can link genes that encode a heterodimeric gene product needed for survival, or can link portions of a replicon.

10 **Site-specific recombinase:** is a type of recombinase which typically has at least the following four activities (or combinations thereof): (1) recognition of one or two specific nucleic acid sequences; (2) cleavage of said sequence or sequences; (3) topoisomerase activity involved in strand exchange; and (4) ligase activity to reseal the cleaved strands of nucleic acid. See Sauer, B., *Current Opinions in Biotechnology* 5:521-527 (1994). Conservative site-specific recombination is distinguished from homologous recombination and transposition by a high degree of specificity for both partners. The strand exchange mechanism involves the cleavage and rejoicing of specific DNA sequences in the absence of DNA synthesis (Landy, A. (1989) *Ann. Rev. Biochem.* 58:913-949).

15 **Subcloning vector:** is a cloning vector comprising a circular or linear nucleic acid molecule which includes preferably an appropriate replicon. In the present invention, the subcloning vector (segment **D** in Figure 1) can also contain functional and/or regulatory elements that are desired to be incorporated into the final product to act upon or with the cloned DNA Insert (segment **A** in Figure 1). The subcloning vector can also contain a Selectable marker (preferably DNA).

20 **Template:** refers to double stranded or single stranded nucleic acid molecules which are to be amplified, synthesized or sequenced. In the case of double stranded molecules, denaturation of its strands to form a first and a second strand is preferably performed before these molecules will be amplified, synthesized or sequenced, or the double stranded molecule may be used directly as a template. For single stranded templates, a primer complementary to a portion of the template is hybridized under appropriate conditions and one or

more polypeptides having polymerase activity (e.g. DNA polymerases and/or reverse transcriptases) may then synthesize a nucleic acid molecule complementary to all or a portion of said template. Alternatively, for double stranded templates, one or more promoters may be used in combination with one or more polymerases to make nucleic acid molecules complementary to all or a portion of the template. The newly synthesized molecules, according to the invention, may be equal or shorter in length than the original template. Additionally, a population of nucleic acid templates may be used during synthesis or amplification to produce a population of nucleic acid molecules typically representative of the original template population.

**Vector:** is a nucleic acid molecule (preferably DNA) that provides a useful biological or biochemical property to an Insert. Examples include plasmids, phages, autonomously replicating sequences (ARS), centromeres, and other sequences which are able to replicate or be replicated *in vitro* or in a host cell, or to convey a desired nucleic acid segment to a desired location within a host cell. A Vector can have one or more restriction endonuclease recognition sites at which the sequences can be cut in a determinable fashion without loss of an essential biological function of the vector, and into which a nucleic acid fragment can be spliced in order to bring about its replication and cloning. Vectors can further provide primer sites, *e.g.*, for PCR, transcriptional and/or translational initiation and/or regulation sites, recombinational signals, replicons, Selectable markers, *etc.* Clearly, methods of inserting a desired nucleic acid fragment which do not require the use of homologous recombination, transpositions or restriction enzymes (such as, but not limited to, UDG cloning of PCR fragments (U.S. Patent No. 5,334,575, entirely incorporated herein by reference), T:A cloning, and the like) can also be applied to clone a fragment into a cloning vector to be used according to the present invention. The cloning vector can further contain one or more selectable markers suitable for use in the identification of cells transformed with the cloning vector.

**Vector Donor:** is one of the two parental nucleic acid molecules (*e.g.* RNA or DNA) of the present invention which carries the DNA segments

comprising the DNA vector which is to become part of the desired Product. The Vector Donor comprises a subcloning vector **D** (or it can be called the cloning Vector if the Insert Donor does not already contain a cloning vector) and a vector if the Insert Donor does not already contain a cloning vector) and a segment **C** flanked by recombination sites (see Figure 1). Segments **C** and/or **D** can contain elements that contribute to selection for the desired Product daughter molecule, as described above for selection schemes. The recombination signals can be the same or different, and can be acted upon by the same or different recombinases. In addition, the Vector Donor can be linear or circular.

Other terms used in the fields of recombinant DNA technology and molecular and cell biology as used herein will be generally understood by one of ordinary skill in the applicable arts.

#### Recombination Schemes

One general scheme for an *in vitro* or *in vivo* method of the invention is shown in Figure 1, where the Insert Donor and the Vector Donor can be either circular or linear DNA, but is shown as circular. Vector **D** is exchanged for the original cloning vector **B**. The Insert Donor need not comprise a vector. The method of the invention allows the Inserts **A** to be transferred into any number of vectors. According to the invention, the Inserts may be transferred to a particular Vector or may be transferred to a number of vectors in one step. Additionally, the Inserts may be transferred to any number of vectors sequentially, for example, by using the Product DNA molecule as the Insert Donor in combination with a different Vector Donor. The nucleic acid molecule of interest may be transferred into a new vector thereby producing a new Product DNA molecule. The new Product DNA molecule may then be used as starting material to transfer the nucleic acid molecule of interest into a new vector. Such sequential transfers can be performed a number of times in any number of different vectors. Thus the invention allows for cloning or subcloning nucleic acid molecules and because of the ease and simplicity, these methods are particularly suited for high throughput applications. In accordance with the invention, it is desirable to select for the daughter molecule containing elements **A** and **D** and against other molecules,

including one or more Cointegrate(s). The square and circle are different sets of recombination sites (e.g., *lox* sites or *att* sites). Segment **A** or **D** can contain at least one Selection Marker, expression signals, origins of replication, or specialized functions for detecting, selecting, expressing, mapping or sequencing DNA, where **D** is used in this example. This scheme can also be reversed according to the present invention, as described herein. The resulting product of the reverse reaction (e.g. the Insert Donor) may then be used in combination with one or a number of vectors to produce new product molecules in which the Inserts are contained by any number of vectors.

Examples of desired DNA segments that can be part of Element **A** or **D** include, but are not limited to, PCR products, large DNA segments, genomic clones or fragments, cDNA clones or fragments, functional elements, *etc.*, and genes or partial genes, which encode useful nucleic acids or proteins. Moreover, the recombinational cloning of the present invention can be used to make *ex vivo* and *in vivo* gene transfer vehicles for protein expression (native or fusion proteins) and/or gene therapy.

In Figure 1, the scheme provides the desired Product as containing **A** and Vector **D**, as follows. The Insert Donor (containing **A** and **B**) is first recombined at the square recombination sites by recombination proteins, with the Vector Donor (containing **C** and **D**), to form a Co-integrate having each of **A-D-C-B**. Next, recombination occurs at the circle recombination sites to form Product DNA (**A** and **D**) and Byproduct DNA **C** and **B**). However, if desired, two or more different Co-integrates can be formed to generate two or more Products.

Recombinational cloning using nucleic acid molecules comprising engineered recombination sites, and the materials and methods by which this technique may be accomplished, have been described in detail in U.S. Application Nos. 08/486,139, filed June 7, 1995 (now abandoned), 08/663,002, filed June 7, 1996 (now U.S. Patent No. 5,888,732), 09/005,476, filed January 12, 1998, 60/065,930, filed October 24, 1997, 09/177,387, filed October 23, 1998, 60/122,389, filed March 2, 1999, 60/122,392, filed March 22, 1999, 60/126,049, filed March 23, 1999, and 60/136,744, filed

May 28, 1999. The disclosures of all of the above-referenced patent applications are incorporated herein by reference in their entireties for their relevant teachings.

### Compositions

By the present invention, compositions are provided that may be used in recombinational cloning of nucleic acid molecules or segments thereof. Compositions of the invention may comprise mixtures of at least one ribosomal protein and at least one recombination protein, suitable for use in the recombinational cloning of nucleic acid molecules. The compositions of the invention may comprise two or more, three or more, four or more, five or more, etc., ribosomal proteins, recombination proteins, or combinations thereof. In related embodiments, the compositions may further comprise one or more additional components, such as one or more nucleic acid molecules (including, but not limited to, one or more Insert Donor molecules, one or more Vector Donor molecules, one or more cointegrate molecules, one or more Product molecules and one or more Byproduct molecules), one or more buffer salts, and/or other reagents which may be used in recombinational cloning of nucleic acid molecules. In related aspects, the ribosomal proteins, recombination proteins, and/or compositions of the invention may contain one or more stabilizing compounds (e.g., glycerol, serum albumin or gelatin) that are traditionally included in stock reagent solutions. Suitable amounts of such stabilizing compounds will be familiar to one of ordinary skill in the art, or may be easily determined using only routine experimentation. For example, glycerol may be used in the compositions of the invention at a concentration (vol/vol) of about 5%-75%, about 10%-65%, about 15%-60%, about 20%-55%, about 25%-50%, or about 50%. In an additional related aspect, the invention provides these compositions in ready-to-use concentrations, obviating the time-consuming dilution and pre-mixing steps necessary with previously available solutions.

*Ribosomal Proteins.* The one or more ribosomal proteins used in the present compositions may be basic ribosomal proteins. By a "basic" ribosomal protein is meant a ribosomal protein that comprises a relatively high percentage

(i.e., ranging from about 15-50%) of basic amino acid residues, particularly arginine and lysine. The ribosomal proteins used in the compositions and methods of the invention preferably are no larger than about 14 kilodaltons (kD) in size, and more preferably are about 5 kD to about 14 kD, about 6 kD to about 13 kD, about 7 kD to about 12 kD, or about 8 kD to about 12 kD, in size.

According to the invention, the one or more ribosomal proteins may be one or more prokaryotic ribosomal proteins (e.g., one or more bacterial ribosomal proteins) or one or more eukaryotic ribosomal proteins, e.g., one or more ribosomal proteins of animals (such as mammals (including humans), fish, birds, reptiles, amphibians, monotremes, and the like), fungi, plants, and the like. In certain compositions, the ribosomal proteins may be one or more prokaryotic ribosomal proteins, particularly one or more ribosomal proteins obtained from bacteria including, but not limited to, those of the genera *Escherichia*, *Serratia*, *Salmonella*, *Pseudomonas*, *Bacillus*, *Streptomyces*, *Staphylococcus*, *Streptococcus*, or other gram positive or gram negative bacteria.

In particularly preferred compositions of the invention, the ribosomal proteins may be one or more *Escherichia coli* ribosomal proteins. Particularly preferred such *E. coli* ribosomal proteins for use in the compositions and methods of the invention include, but are not limited to, S10, S14, S15, S16, S17, S18, S19, S20, S21, L21, L23, L24, L25, L27, L28, L29, L30, L31, L32, L33 and L34. Most preferred *E. coli* ribosomal proteins for use in the compositions and methods of the invention are S20, L27 and S15. Corresponding ribosomal proteins from other sources, including prokaryotic or eukaryotic sources, may be used in accordance with the invention. Such corresponding ribosomal proteins preferably correspond (in structure, size, biochemistry, and/or function) to the *E. coli* ribosomal proteins described herein.

Sources and methods for production and isolation of ribosomal proteins, particularly prokaryotic ribosomal proteins, are described in detail in Example 1 below. In addition, information on sources and isolation of prokaryotic and eukaryotic ribosomal proteins may be found in *Ann. Rev. Biochem.* 51:155 (1982); *Ann. Rev. Biochem.* 52:35 (1983); *Ann. Rev. Biochem.* 53:75 (1984); *Ann.*

*Rev. Biochem.* 54:507 (1985); *Ann. Rev. Biochem.* 66:679 (1997); and Bruckner and Cox, *Nucl. Acids Res.* 17(8):3145-3161 (1989).

The amount of one or more ribosomal proteins which is optimal for use in the compositions and methods of the present invention to drive the recombination reaction can be determined using known assays. Specifically, a titration assay may be used to determine the appropriate amount of a purified ribosomal protein, or the appropriate amount of an extract. Such assays are described in detail in the Examples below. In certain embodiments, for example, the compositions may comprise an effective amount of the *E. coli* ribosomal proteins S20 or S15, for example at a concentration range of about 1 ng to about 2500 ng, about 2 ng to about 2000 ng, about 5 ng to about 1500 ng, about 10 ng to about 1500 ng, about 25 ng to about 1500 ng, about 50 ng to about 1500 ng, about 100 ng to about 1500 ng, about 250 ng to about 1500 ng, about 300 ng to about 1500 ng, about 500 ng to about 1500 ng, about 500 ng to about 1250 ng, or about 625 ng to about 1250 ng. In other embodiments, the compositions may comprise the *E. coli* ribosomal protein L27, at a concentration of, for example, about 1,000 ng to about 50,000 ng, about 2,000 ng to about 40,000 ng, about 5,000 ng to about 30,000 ng, about 10,000 ng to about 25,000 ng, about 10,000 ng to about 20,000 ng, or about 10,000 ng. Of course, other concentration ranges for S20, S15, or L27, or other suitable prokaryotic or eukaryotic ribosomal proteins that may be used in the present compositions, may be determined by one of ordinary skill without undue experimentation by carrying out a titration assay as noted above and as described in detail in the Examples below.

*Recombination Proteins.* In the compositions and methods of the present invention, the exchange of DNA segments is achieved by the use of recombination proteins, including recombinases and associated co-factors and proteins. The one or more recombination proteins for use in the compositions may be any recombination protein, including any prokaryotic or eukaryotic recombination protein, that is suitable for use in recombinational cloning of nucleic acid molecules. Examples of such recombination proteins include, but are not limited to:

**Cre:** A prokaryotic recombination protein from bacteriophage P1 (Abremski and Hoess, *J. Biol. Chem.* 259(3):1509-1514 (1984)) catalyzes the exchange (i.e., causes recombination) between 34 bp DNA sequences called *loxP* (locus of crossover) sites (See Hoess *et al.*, *Nucl. Acids Res.* 14(5):2287 (1986)). Cre is available commercially (Novagen, Catalog No. 69247-1). Recombination mediated by Cre is freely reversible. From thermodynamic considerations it is not surprising that Cre-mediated integration (recombination between two molecules to form one molecule) is much less efficient than Cre-mediated excision (recombination between two *loxP* sites in the same molecule to form two daughter molecules). Cre works in simple buffers with either magnesium or spermidine as a cofactor, as is well known in the art. The DNA substrates can be either linear or supercoiled. A number of mutant *loxP* sites have been described (Hoess *et al.*, *supra*). One of these, *loxP* 511, recombines with another *loxP* 511 site, but will not recombine with a *loxP* site.

**Integrase:** A prokaryotic recombination protein from bacteriophage lambda that mediates the integration of the lambda genome into the *E. coli* chromosome. The bacteriophage  $\lambda$  Int recombinational proteins promote recombination between its substrate *att* sites as part of the formation or induction of a lysogenic state. Reversibility of the recombination reactions results from two independent pathways for integrative and excisive recombination. Each pathway uses a unique, but overlapping, set of the 15 protein binding sites that comprise *att* site DNAs. Cooperative and competitive interactions involving four proteins (Int, Xis, IHF and FIS) determine the direction of recombination.

Integrative recombination involves the Int and IHF proteins and sites *attP* (240 bp) and *attB* (25 bp). Recombination results in the formation of two new sites: *attL* and *attR*. Excisive recombination requires Int, IHF, and Xis, and sites *attL* and *attR* to generate *attP* and *attB*. Under certain conditions, FIS stimulates excisive recombination. In addition to these normal reactions, it should be appreciated that *attP* and *attB*, when placed on the same molecule, can promote excisive recombination to generate two excision products, one with *attL* and one with *attR*. Similarly, intermolecular recombination between molecules containing

attL and attR, in the presence of Int, IHF and Xis, can result in *integrative* recombination and the generation of attP and attB. Hence, by flanking DNA segments with appropriate combinations of engineered att sites, in the presence of the appropriate recombination proteins, one can direct excisive or integrative recombination, as reverse reactions of each other.

Each of the att sites contains a 15 bp core sequence; individual sequence elements of functional significance lie within, outside, and across the boundaries of this common core (Landy, A., *Ann. Rev. Biochem.* 58:913 (1989)). Efficient recombination between the various att sites requires that the sequence of the central common region be identical between the recombining partners, however, the exact sequence is now found to be modifiable. Consequently, derivatives of the att site with changes within the core are now discovered to recombine as least as efficiently as the native core sequences.

Integrase acts to recombine the attP site on bacteriophage lambda (about 240 bp) with the attB site on the *E. coli* genome (about 25 bp) (Weisberg, R.A. and Landy, A. in *Lambda II*, p. 211 (1983), Cold Spring Harbor Laboratory), to produce the integrated lambda genome flanked by attL (about 100 bp) and attR (about 160 bp) sites. In the absence of Xis (see below), this reaction is essentially irreversible. The integration reaction mediated by integrase and IHF works *in vitro*, with simple buffer containing spermidine. Integrase can be obtained as described by Nash, H.A., *Methods of Enzymology* 100:210-216 (1983). IHF can be obtained as described by Filutowicz, M., *et al.*, *Gene* 147:149-150 (1994).

Numerous recombination systems from various organisms can also be used, based on the teaching and guidance provided herein. See, e.g., Hoess *et al.*, *Nucleic Acids Research* 14(6):2287 (1986); Abramski *et al.*, *J. Biol. Chem.* 261(1):391 (1986); Campbell, *J. Bacteriol.* 174(23):7495 (1992); Qian *et al.*, *J. Biol. Chem.* 267(11):7794 (1992); Araki *et al.*, *J. Mol. Biol.* 225(1):25 (1992)). Many of these belong to the integrase family of recombinases (Argos *et al.* *EMBO J.* 5:433-440 (1986)). Perhaps the best studied of these are the Integrase/att system from bacteriophage  $\lambda$  (Landy, A. (1993) *Current Opinions in Genetics and Devel.* 3:699-707), the Cre/loxP system from bacteriophage P1

(Hoess and Abremski (1990) In *Nucleic Acids and Molecular Biology*, vol. 4. Eds.: Eckstein and Lilley, Berlin-Heidelberg: Springer-Verlag; pp. 90-109), and the FLP/FRT system from the *Saccharomyces cerevisiae* 2  $\mu$  circle plasmid (Broach *et al. Cell* 29:227-234 (1982)).

5 Members of the resolvase (Res) family of site-specific recombinases (e.g.,  $\gamma\delta$ , Tn3 resolvase, Hin, Gin, and Cin) are also known, and may be used in accordance with the present invention. Members of this highly related family of recombinases are typically constrained to intramolecular reactions (e.g., inversions and excisions) and can require host-encoded factors. Mutants have  
10 been isolated that relieve some of the requirements for host factors (Maeser and Kahnmann (1991) *Mol. Gen. Genet.* 230:170-176), as well as some of the constraints of intramolecular recombination.

Other site-specific recombinases similar to  $\lambda$  Int and similar to P1 Cre can be substituted for Int and Cre. Such recombinases are known. In many cases the purification of such other recombinases has been described in the art. In cases when they are not known, cell extracts can be used or the enzymes can be partially purified using procedures described for Cre and Int.

While Cre and Int are described in detail for reasons of example, many related recombination systems and proteins exist and their application to the described invention is also provided according to the present invention. The integrase family of site-specific recombinases can be used to provide alternative recombination proteins and recombination sites for the present invention, as site-specific recombination proteins encoded by, for example bacteriophage lambda, phi 80, P22, P2, 186, P4 and P1. This group of recombination proteins, which may be used in the present compositions and methods, exhibits an unexpectedly large diversity of sequences. Despite this diversity, all of these recombinases can be aligned in their C-terminal halves. A 40-residue region near the C terminus is particularly well conserved in all the proteins and is homologous to a region near the C terminus of the yeast 2  $\mu$  plasmid FLP recombination protein. Three positions are perfectly conserved within this family: histidine, arginine and tyrosine are found at respective alignment positions 396, 399 and

433 within the well-conserved C-terminal region. These residues contribute to the active site of this family of recombinases, and suggest that tyrosine-433 forms a transient covalent linkage to DNA during strand cleavage and rejoicing. See, e.g., Argos, P. et al., *EMBO J.* 5:433-40 (1986).

5 The recombinases of some transposons, such as those of conjugative transposons (e.g., Tn916) (Scott and Churchward, 1995. *Ann Rev Microbiol* 49:367; Taylor and Churchward, 1997. *J Bacteriol* 179:1837), may also be used in the compositions and methods of the invention. These transposon recombinases belong to the integrase family of recombinases and in some cases show strong preferences for specific integration sites (Ike et al 1992. *J Bacteriol* 174:1801; Trieu-Cuot et al, 1993. *Mol. Microbiol* 8:179).

10 Alternatively, IS231 and other *Bacillus thuringiensis* transposable elements could be used in accordance with the present invention as recombination proteins and recombination sites. *Bacillus thuringiensis* is an entomopathogenic bacterium whose toxicity is due to the presence in the sporangia of delta-endotoxin crystals active against agricultural pests and vectors of human and animal diseases. Most of the genes coding for these toxin proteins are plasmid-borne and are generally structurally associated with insertion sequences (IS231, IS232, IS240, ISBT1 and ISBT2) and transposons (Tn4430 and Tn5401). Several of these mobile elements have been shown to be active and participate in the crystal gene mobility, thereby contributing to the variation of bacterial toxicity.

25 Structural analysis of the iso-IS231 elements indicates that they are related to IS1151 from *Clostridium perfringens* and distantly related to IS4 and IS186 from *Escherichia coli*. Like the other IS4 family members, they contain a conserved transposase-integrase motif found in other IS families and retroviruses. Moreover, functional data gathered from IS231A in *Escherichia coli* indicate a non-replicative mode of transposition, with a preference for specific targets. Similar results were also obtained in *Bacillus subtilis* and *B. thuringiensis*. See, e.g., Mahillon, J. et al., *Genetica* 93:13-26 (1994); Campbell, J. *Bacteriol.* 7495-7499 (1992).

5 An unrelated family of recombinases, the transposases, have also been used to transfer genetic information between replicons, and may therefore be used as recombination proteins in accordance with the invention. Transposons are structurally variable, being described as simple or compound, but typically encode the recombinase gene flanked by DNA sequences organized in inverted orientations. Integration of transposons can be random or highly specific. Representatives such as Tn7, which are highly site-specific, have been applied to the efficient movement of DNA segments between replicons (Lucklow et al. 1993. *J. Virol* 67:4566-4579).

10 A related element, the integron, are also translocatable-promoting movement of drug resistance cassettes from one replicon to another. Often these 0 elements are defective transposon derivatives. Transposon Tn21 contains a class I 5 integron called In2. The integrase (IntI1) from In2 is common to all integrons in this class and mediates recombination between two 59-bp elements or between a 59-bp element and an attI site that can lead to insertion into a recipient integron. 15 The integrase also catalyzes excisive recombination. (Hall, 1997. Ciba Found Symp 207:192; Francia et al., 1997. *J Bacteriol* 179:4419).

20 Group II introns are mobile genetic elements encoding a catalytic RNA and protein. The protein component possesses reverse transcriptase, maturase and an endonuclease activity, while the RNA possesses endonuclease activity and determines the sequence of the target site into which the intron integrates. By 25 modifying portions of the RNA sequence, the integration sites into which the element integrates can be defined. Foreign DNA sequences can be incorporated between the ends of the intron, allowing targeting to specific sites. This process, termed retrohoming, occurs via a DNA:RNA intermediate, which is copied into cDNA and ultimately into double stranded DNA (Matsuura et al., *Genes and Dev* 1997; Guo et al, *EMBO J*, 1997). Numerous intron-encoded homing endonucleases have been identified (Belfort and Roberts, 1997. *NAR* 25:3379). 30 Such systems can be easily adopted for application to the subcloning methods described herein.

In addition, other suitable recombination proteins are described in detail in U.S. Application Nos. 08/486,139, filed June 7, 1995 (now abandoned), 08/663,002, filed June 7, 1996 (now U.S. Patent No. 5,888,732), 09/005,476, filed January 12, 1998, 60/065,930, filed October 24, 1997, 09/177,387, filed October 23, 1998, 60/122,389, filed March 2, 1999, 60/122,392, filed March 22, 1999, 60/126,049, filed March 23, 1999, and 60/136,744, filed May 28, 1999, the disclosures of all of which are incorporated herein by reference in their entireties for their relevant teachings. Hence, in preferred compositions of the invention, the recombination protein may be selected from the group consisting of Int, Cre, Res, Xis, FLP, IHF and HU, and may be a site-specific recombination protein. Particularly preferred for use in the present compositions is Int.

The amount of recombination protein which is optimal for use in the compositions and methods of the present invention to drive the recombination reaction can be determined using known assays. Specifically, a titration assay may be used to determine the appropriate amount of a purified recombination protein, or the appropriate amount of an extract. Such assays are described in detail in the Examples below. In certain preferred compositions of the invention, for example, the compositions may comprise an effective amount of  $\lambda$  Int, for example at a concentration range of about 1 ng to about 500 ng, about 2 ng to about 250 ng, about 5 ng to about 200 ng, about 10 ng to about 200 ng, about 25 ng to about 200 ng, about 50 ng to about 200 ng, or about 100 ng to about 200 ng. In addition, the compositions may comprise one or more additional recombination proteins; for example, a composition of the invention may comprise  $\lambda$  Int at the above-indicated concentration ranges, and HU protein and/or IHF protein at concentration ranges of about 1 ng to about 50 ng, about 2 ng to about 25 ng, about 5 ng to about 20 ng, about 5 ng to about 15 ng, or about 5 ng to about 10 ng. Of course, other concentration ranges for  $\lambda$  Int or other recombination proteins that may be used in the present compositions may be determined by one of ordinary skill, without undue experimentation, by carrying

out a titration assay as noted above and as described in detail in the Examples below.

### Recombinational Cloning Methods

The above-described compositions of the invention are suitable for use in recombination cloning methods that are provided by the present invention. Recombinational cloning using nucleic acid molecules comprising engineered recombination sites, and the materials and methods by which this technique may be accomplished, have been described in detail in U.S. Application Nos. 08/486,139, filed June 7, 1995 (now abandoned), 08/663,002, filed June 7, 1996 (now U.S. Patent No. 5,888,732), 09/005,476, filed January 12, 1998, 60/065,930, filed October 24, 1997, 09/177,387, filed October 23, 1998, 60/122,389, filed March 2, 1999, 60/122,392, filed March 22, 1999, 60/126,049, filed March 23, 1999, and 60/136,744, filed May 28, 1999. The disclosures of all of the above-referenced patent applications are incorporated herein by reference in their entireties for their relevant teachings.

In one such aspect, the invention relates to such methods comprising:

- (a) combining *in vitro* or *in vivo*
  - (i) one or more Insert Donor molecules comprising one or more desired nucleic acid segments flanked by at least two recombination sites, wherein the recombination sites do not substantially recombine with each other;
  - (ii) one or more Vector Donor molecules comprising at least two recombination sites, wherein the recombination sites do not substantially recombine with each other;
  - (iii) at least one recombination protein; and
  - (iv) at least one ribosomal protein;
- (b) incubating the combination formed in step (a) under conditions sufficient to transfer one or more of the desired segments into one or more of the Vector Donor molecules, thereby producing one or more desired Product nucleic acid molecules;

and optionally:

5

- (c) combining *in vitro* or *in vivo*
  - (i) one or more of the Product molecules comprising the desired segments flanked by two or more recombination sites, wherein the recombination sites do not substantially recombine with each other;
  - (ii) one or more different Vector Donor molecules comprising two or more recombination sites, wherein the recombination sites do not substantially recombine with each other;
  - (iii) at least one recombination protein; and
  - (iv) at least one ribosomal protein; and
- (d) incubating the combination formed in step (c) under conditions sufficient to transfer one or more of the desired segments into one or more different Vector Donor molecules, thereby producing one or more different Product molecules.

10

15

The invention also relates to such methods which further comprise incubating the different Product molecules with one or more different Vector Donor molecules under conditions sufficient to transfer one or more of the desired segments into the different Vector Donor molecules.

20

In a related aspect, the invention relates to methods of cloning or subcloning one or more desired nucleic acid molecules by recombinational cloning comprising:

25

- a) combining *in vitro* or *in vivo*
  - i) one or more Insert Donor molecules comprising one or more nucleic acid segments flanked by two or more recombination sites, wherein the recombination sites do not substantially recombine with each other;
  - ii) two or more different Vector Donor molecules comprising two or more recombination sites, wherein the

30

recombination sites do not substantially recombine with each other;

iii) at least one recombination protein; and

iv) at least one ribosomal protein; and

5 b) incubating the combination formed in step (a) under conditions sufficient to transfer one or more of the desired segments into the different Vector Donor molecules, thereby producing two or more different Product molecules.

In another related aspect, the invention relates to methods for recombinational cloning of one or more desired nucleic acid molecules comprising

10 (a) mixing one or more desired nucleic acid molecules with one or more vectors and with one or more of the compositions of the invention; and

(b) incubating the mixture under conditions sufficient to transfer the one or more desired nucleic acid molecules into one or more of the vectors.

In another related aspect, the invention relates to methods for enhancement of recombinational cloning of nucleic acid molecules, comprising contacting one or more nucleic acid molecules with one or more ribosomal proteins and one or more recombination proteins, or with one or more compositions of the invention, under conditions favoring the recombinational cloning of the one or more nucleic acid molecules.

25 According to the invention, the one or more ribosomal proteins used in these methods may be one or more prokaryotic or eukaryotic ribosomal proteins, such as those described herein. Similarly, the one or more recombination proteins may be one or more prokaryotic or eukaryotic recombination proteins such as those described herein.

In another related aspect, the invention relates to methods of cloning or subcloning one or more desired nucleic acid molecules by recombinational cloning comprising:

30 (a) combining *in vitro* or *in vivo*

5

- (i) one or more Insert Donor molecules comprising one or more desired nucleic acid segments flanked by at least two recombination sites, wherein the recombination sites do not substantially recombine with each other;
- (ii) one or more Vector Donor molecules comprising at least two recombination sites, wherein the recombination sites do not substantially recombine with each other; and
- (iii) one or more of the compositions of the invention;

10

- (b) incubating the combination formed in step (a) under conditions sufficient to transfer one or more of the desired segments into one or more of the Vector Donor molecules, thereby producing one or more desired Product nucleic acid molecules;

and optionally:

20

- (c) combining *in vitro* or *in vivo*
  - (i) one or more of the Product molecules comprising the desired segments flanked by two or more recombination sites, wherein the recombination sites do not substantially recombine with each other;
  - (ii) one or more different Vector Donor molecules comprising two or more recombination sites, wherein the recombination sites do not substantially recombine with each other; and
  - (iii) one or more of the compositions of the invention; and
- (d) incubating the combination formed in step (c) under conditions sufficient to transfer one or more of the desired segments into one or more different Vector Donor molecules, thereby producing one or more different Product molecules.

25

In another related aspect, the invention relates to methods of cloning or subcloning one or more desired nucleic acid molecules by recombinational cloning comprising:

30

- a) combining *in vitro* or *in vivo*

- i) one or more Insert Donor molecules comprising one or more nucleic acid segments flanked by two or more recombination sites, wherein the recombination sites do not substantially recombine with each other;
- ii) two or more different Vector Donor molecules comprising two or more recombination sites, wherein the recombination sites do not substantially recombine with each other; and
- iii) one or more of the compositions of the invention; and

b) incubating the combination formed in step (a) under conditions sufficient to transfer one or more of the desired segments into the different Vector Donor molecules, thereby producing two or more different Product molecules.

According to the invention, the Insert Donor molecules for use in the compositions and methods of the invention may be derived from genomic DNA or cDNA, or may be produced by chemical synthesis methods. In a related aspect, the Insert Donor molecules may comprise one or more vectors.

The Vector Donor molecules for use in the compositions and methods of the invention may optionally comprise at least one Selectable marker, which allows for the selection of host cells comprising the Product molecules comprising the Selectable markers contributed by the Vector Donor molecules during the recombination reaction. According to this aspect of the invention, the Selectable Marker may be an antibiotic resistance gene, a tRNA gene, an auxotrophic marker, a toxic gene, a phenotypic marker, an antisense oligonucleotide, a restriction endonuclease, a restriction endonuclease cleavage site, an enzyme cleavage site, a protein binding site, and a sequence complementary to a PCR primer sequence. In a related aspect, the Vector Donor molecules may comprise one or more eukaryotic vectors or one or more prokaryotic vectors. Eukaryotic vectors suitable for use in this aspect of the invention may comprise, for example, vectors which propagate and/or replicate in yeast cells, plant cells, fish cells, eukaryotic cells, mammalian cells, and/or

insect cells, while suitable prokaryotic vectors may comprise, for example, vectors which propagate and/or replicate in bacteria of the genera *Escherichia* (most particularly *E. coli*), *Salmonella*, *Bacillus*, *Streptomyces* or *Pseudomonas*.

5 The compositions and methods described herein are suitable for use in recombination cloning according to the present invention. However, wild-type recombination sites that are contained in the Insert Donor and/or Vector Donor DNA molecules may contain sequences that reduce the efficiency or specificity of recombination reactions or the function of the Product molecules as applied in methods of the present invention. For example, multiple stop codons in attB, attR, attP, attL and loxP recombination sites occur in multiple reading frames on both strands, so translation efficiencies are reduced, e.g., where the coding sequence must cross the recombination sites, (only one reading frame is available on each strand of loxP and attB sites) or impossible (in attP, attR or attL).

10 Accordingly, DNA molecules comprising one or more engineered recombination sites are preferably used in the methods of the present invention, to overcome these problems. For example, att sites can be engineered to have one or multiple mutations to enhance specificity or efficiency of the recombination reaction and the properties of Product DNAs (e.g., att1, att2, and att3 sites); to decrease reverse reaction (e.g., removing P1 and H1 from attR). The testing of these mutants determines which mutants yield sufficient recombinational activity to be suitable for recombination subcloning according to the present invention. Hence, in addition to the one or more ribosomal proteins and one or more recombination proteins described herein, the compositions of the invention may further comprise one or more nucleic acid molecules including, but not limited to, one or more Insert Donor molecules, one or more Vector Donor molecules, one or more cointegrate molecules, one or more Product molecules and one or more Byproduct molecules, any or all of which may contain engineered or mutant recombination sites.

25 30 Mutations can be introduced into recombination sites for enhancing site specific recombination. The production of DNA molecules comprising one or more mutated engineered recombination sites, which molecules may be used as

Insert Donor or Vector Donor molecules in the recombinational cloning methods of the present invention, is described in detail in Application Nos. 08/486,139, filed June 7, 1995 (now abandoned), 08/663,002, filed June 7, 1996 (now U.S. Patent No. 5,888,732), 09/005,476, filed January 12, 1998, 60/065,930, filed October 24, 1997, 09/177,387, filed October 23, 1998, 60/122,389, filed March 2, 1999, 60/122,392, filed March 22, 1999, 60/126,049, filed March 23, 1999, and 60/136,744, filed May 28, 1999, the disclosures of all of which applications are incorporated herein by reference in their entireties. Particularly preferred for use in the compositions and methods of the present invention are nucleic acid molecules comprising at least one DNA segment having at least two engineered recombination sites flanking a Selectable marker and/or a desired DNA segment, wherein at least one of the recombination sites comprises a core region having at least one engineered mutation that enhances recombination *in vitro* in the formation of a Cointegrate DNA or a Product DNA.

In accordance with the invention, any vector may be used to construct the Vector Donors used in the methods of the invention. In particular, vectors known in the art and those commercially available (and variants or derivatives thereof) may in accordance with the invention be engineered to include one or more recombination sites for use in the methods of the invention. Such vectors may be obtained from, for example, Vector Laboratories Inc., InVitrogen, Promega, Novagen, NEB, Clontech, Boehringer Mannheim, Pharmacia, EpiCenter, OriGenes Technologies Inc., Stratagene, Perkin Elmer, Pharmingen, Life Technologies, Inc., and Research Genetics. Such vectors may then for example be used for cloning or subcloning nucleic acid molecules of interest. General classes of vectors of particular interest include prokaryotic and/or eukaryotic cloning vectors, expression vectors, fusion vectors, two-hybrid or reverse two-hybrid vectors, shuttle vectors for use in different hosts, mutagenesis vectors, transcription vectors, vectors for receiving large inserts and the like. Particularly preferred vectors (and mutants, derivatives, or variants thereof) that may be used to construct the Vector Donors used in the methods of the invention are described in detail in U.S. Application Nos. 08/486,139, filed June 7, 1995 (now

abandoned), 08/663,002, filed June 7, 1996 (now U.S. Patent No. 5,888,732), 09/005,476, filed January 12, 1998, 60/065,930, filed October 24, 1997, 09/177,387, filed October 23, 1998, 60/122,389, filed March 2, 1999, 60/122,392, filed March 22, 1999, 60/126,049, filed March 23, 1999, and 60/136,744, filed May 28, 1999, the disclosures of all of which applications are incorporated herein by reference in their entireties.

5

#### **DNA Molecules, Vectors and Host Cells**

The invention also relates generally to DNA molecules produced by the methods of the invention, particularly to such DNA molecules which are isolated DNA molecules. Methods for the isolation of DNA molecules produced by the methods of the invention will be familiar to one of ordinary skill in the art, and are described generally in U.S. Application Nos. 08/486,139, filed June 7, 1995 (now abandoned), 08/663,002, filed June 7, 1996 (now U.S. Patent No. 5,888,732), 09/005,476, filed January 12, 1998, 60/065,930, filed October 24, 1997, 09/177,387, filed October 23, 1998, 60/122,389, filed March 2, 1999, 60/122,392, filed March 22, 1999, 60/126,049, filed March 23, 1999, and 60/136,744, filed May 28, 1999, the disclosures of which are incorporated herein by reference in their entireties. In addition, the isolated DNA molecules of the invention may be inserted into standard nucleotide vectors suitable for transfection or transformation of a variety of prokaryotic (bacterial) or eukaryotic (yeast, plant or animal including human and other mammalian) host cells. Vectors suitable for these purposes, and methods for insertion of DNA fragments therein, will be well-known to one of ordinary skill in the art. Thus, the present invention also relates to vectors comprising such DNA molecules, and to host cells comprising such DNA molecules and/or vectors.

10

116

117

118

119

120

121

122

123

124

125

126

127

128

129

130

#### **Kits**

The invention also relates to kits for use in recombinational cloning of a nucleic acid molecule. Kits according to the present invention may comprise a carrying means being compartmentalized to receive in close confinement therein

one or more containers such as vials, tubes, bottles, ampules and the like. Each of such containers may comprise components or a mixture of components needed to perform recombinational cloning of nucleic acid molecules, particularly according to the methods of the present invention.

5 In one such aspect, the kits of the invention may comprise at least one ribosomal protein and at least one recombination protein. Ribosomal proteins and recombination proteins suitable for use in the kits of the invention include, but are not necessarily limited to, those prokaryotic and eukaryotic ribosomal and recombination proteins described in detail herein. Of course, it is also possible to combine one or more of these components into a single container, such that the kit will contain one or more containers wherein a first container contains at least one ribosomal protein and at least one recombination protein, or wherein a first container contains one or more of the above-described compositions of the invention. Additional kits of the invention may comprise one or more additional containers containing additional components which may be useful in carrying out recombinational cloning of nucleic acid molecules, including, for example, one or more polymerases (such as one or more thermostable DNA polymerases like *Tag*, *Tne*, *Tma*, and the like), one or more polypeptides having reverse transcriptase activity (such as RSV or ASLV reverse transcriptases, particularly those that are substantially reduced in RNase H activity), one or more restriction endonucleases, one or more buffers, one or more detergents, and the like.

#### *Applications*

25 There are a number of applications for the compositions, methods and kits of the present invention. These uses include, but are not limited to, changing vectors, operably linking genes to regulatory genetic sequences (e.g., promoters, enhancers, and the like), constructing genes for fusion proteins, changing copy number, changing replicons, cloning into phages, and cloning, e.g., PCR products (with an *attB* site at one end and a *loxP* site at the other end), genomic DNAs, and cDNAs. Such applications are described in detail, for example, in U.S. Application Nos. 08/486,139, filed June 7, 1995 (now abandoned), 08/663,002,

5 filed June 7, 1996 (now U.S. Patent No. 5,888,732), 09/005,476, filed January 12, 1998, 60/065,930, filed October 24, 1997, 09/177,387, filed October 23, 1998, 60/122,389, filed March 2, 1999, 60/122,392, filed March 22, 1999, 60/126,049, filed March 23, 1999, and 60/136,744, filed May 28, 1999, which was filed on October 23, 1998, the disclosures of all of which applications are incorporated herein by reference in their entireties.

10 It will be understood by one of ordinary skill in the relevant arts that other suitable modifications and adaptations to the methods and applications described herein are readily apparent and may be made without departing from the scope of the invention or any embodiment thereof. Having now described the present invention in detail, the same will be more clearly understood by reference to the following examples, which are included herewith for purposes of illustration only and are not intended to be limiting of the invention.

### *Examples*

15 The present recombinational cloning methods accomplish the exchange of nucleic acid segments to render something useful to the user, such as a change of cloning vectors. These segments must be flanked on both sides by recombination signals that are in the proper orientation with respect to one another. In the examples below the two parental nucleic acid molecules (e.g., plasmids) are called the Insert Donor and the Vector Donor. The Insert Donor contains a segment that will become joined to a new vector contributed by the Vector Donor. The recombination intermediate(s) that contain(s) both starting molecules is called the Cointegrate(s). The second recombination event produces two daughter molecules, called the Product (the desired new clone) and the Byp product.

25

30

### **Buffers**

Various known buffers can be used in the reactions of the present invention. For restriction enzymes, it is advisable to use the buffers recommended by the manufacturer. Alternative buffers can be readily found in the literature or can be devised by those of ordinary skill in the art. One exemplary buffer for lambda integrase is comprised of 50 mM Tris-HCl, at pH 7.5-7.8, 70 mM KCl, 5 mM spermidine, 0.5 mM EDTA, and 0.25 mg/ml bovine serum albumin, and optionally, 10% glycerol. Suitable buffers for other site-specific recombinases which are similar to lambda Int are either known in the art or can be determined empirically by the ordinarily skilled artisan, particularly in light of the above-described buffers.

### **Example 1: Stimulation of Integrase by *E. coli* Ribosomal Proteins**

### **MATERIALS AND METHODS**

**DNA for Recombination Assays.** Plasmid pHN894 (Figure 2), bearing an *attP* site, and plasmid pBB105 (Figure 3), bearing an *attB* site, are described (Kitts, P.A. and Nash, H.A. *J. Mol. Biol.* 204: 95-107 (1988); Nash, H.A. *Methods Enz.* 100: 210-216 (1983)). pBB105 was cut with *Eco*RI before use. Plasmid pHN872 (Figure 4), bearing an *attL* site, and plasmid pHN868 (Figure 5), bearing an *attR* site, are described (Kitts, P.A. and Nash, H.A. *J. Mol. Biol.* 204: 95-107 (1988)). pHN872 was cut with *Sal*I before use. These plasmids were propagated in *E. coli* strain DH10B. To grow cells for preparation of plasmid DNA, the growth medium contained in one liter: 12 g of tryptone, 24 g of yeast extract, 2.3 g of  $\text{KH}_2\text{PO}_4$ , 12.5 g of  $\text{K}_2\text{HPO}_4$ , 0.01% (v/v) PPG antifoam, and appropriate antibiotic. Cells from a glycerol seed were placed in 25 ml of medium containing 100 $\mu$ g/ml ampicillin (pBB105, pHN894, pHN868) or 100  $\mu$ g/ml kanamycin (pHN872) and grown overnight at 37°C. Fifteen ml of overnight culture was added to 1.5 L medium containing 10  $\mu$ g/ml appropriate antibiotic and cells were grown to a  $A_{600}$  of ~ 2.0. Chloramphenicol was then added to a

final concentration of 170  $\mu$ g/ml and growth was continued for 16 hr at 37 °C. Cells were harvested by centrifugation and stored at -70°C. Plasmid DNAs were purified as follows. Frozen cells were thawed on ice and suspended in 7 ml/g cells of 25 mM Tris-HCl (pH 8.0), 10 mM EDTA, and 50 mM glucose (TEG) + 100  $\mu$ g/ml of RNaseA + 1 mg/ml lysozyme. A solution of 1% (w/v) SDS- 0.125 N NaOH at 14 ml/g cells was then added to lyse cells. After 10 minutes on ice, 7.5 M ammonium acetate at 10.5 ml/g cells was added. After 10 minutes on ice, the mixture was centrifuged at 28,000  $\times$  g for 10 minutes and the supernatant was collected. DNA was precipitated by addition of 0.6 volumes of cold isopropanol, and DNA was pelleted by centrifugation at 28,000  $\times$  g for 10 minutes. The DNA pellet was dissolved in 10 mM Tris-HCl (pH 7.5) - 1 mM EDTA ( $T_{10}E_1$ ) + RNase A (100  $\mu$ g/ml) + RNaseT1 (1,200 U/ml). After phenol extraction and ethanol precipitation of the DNA, it was dissolved in  $T_{10}E_1$ . The DNA was dialyzed against 100 volumes of 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 450 mM NaCl ( $T_{10}E_1N_{450}$ ) overnight. The dialyzed DNA was applied to a NACS-37 column (LTI) equilibrated in  $T_{10}E_1N_{450}$ . The column was washed with 10 column volumes of  $T_{10}E_1N_{450}$  and eluted with a 15-column volume linear gradient from 0.45 M to 0.65 M NaCl in  $T_{10}E_1$ . Fractions were analyzed by agarose gel electrophoresis and those containing supercoiled DNA were pooled. The pooled DNA was dialyzed against  $T_{10}E_1$  and stored at -20°C.

Plasmid pEZ13835 (Figure 6; *attP*), pEZC7501 (Figure 7; *attB*), pEZ11104 (Figure 8; *attR*), and pEZC8402 (Figure 9; *attL*) were as shown. pEZC7501 was cut with *Sal*I and pEZC8402 with *Nco*I before use. pEZ13835 and pEZC8402 were propagated in *E. coli* DB2 and the other two in *E. coli* DH5 $\alpha$ . Cells from a glycerol seed were placed in 25 ml of Circlegrow (BIO 101) plus 100  $\mu$ g/ml ampicillin (pEZC7501 and pEZC8402) or plus 100  $\mu$ g/ml kanamycin (pEZ13835 and pEZ11104) and grown overnight at 37 °C. Cells were harvested by centrifugation and stored at -70 °C. Plasmid DNAs were purified using Qiagen Midi products and protocols.

**SDS PAGE.** Tris-Tricine SDS PAGE 16% precast mini gels (Novex) were used to analyze protein samples. The samples were prepared by mixing with

an equal volume of 0.9 M Tris-HCl (pH 8.45), 24% (v/v) glycerol, 8% (w/v) SDS, 0.015% (w/v) Coomassie BlueG, 0.005% (w/v) Phenol Red, and 0.05 M dithiothreitol and boiling for 3 to 5 min. Gels were run at 125 volts in 0.1 M Tris-Tricine (pH 8.3)- 0.1% (w/v) SDS for 90 min. Gels were stained in 50% (v/v) methanol, 10% (v/v) acetic acid, and 1 mg/ml Coomassie Blue R-250 solution followed by destaining in 20% (v/v) methanol, 10% (v/v) acetic acid solution.

5

**Determination of Protein Concentration.** S20, Int, and Xis bind Bradford reagent dye poorly, so that the Bradford procedure was not used to determine protein concentration. Rather, for Int and Xis, protein concentration was estimated by comparison to Coomassie Blue-stained band intensities of a known amount of BenchMark protein standard of a similar size run along with Int or Xis on an SDS gel. For S20, protein concentration was established using an extinction coefficient at 278 nm of  $0.140 \times 10^4 \text{ M}^{-1}\text{cm}^{-1}$  (*Eur. J. Biochem.* 126: 299-309 (1982)).

10

**PCR.** PCR reaction mixtures (50  $\mu\text{l}$ ) contained 22 mM Tris-HCl (pH 8.4), 55 mM KCl, 1.65 mM MgCl<sub>2</sub>, 200  $\mu\text{M}$  each of dATP, dCTP, dTTP, and dGTP, 1  $\mu\text{M}$  of each primer, 300 ng of DNA template, and 1.1 units of Taq DNA polymerase. Initial template denaturation was at 95 °C for 5 minutes.

15

**Purification of IHF.** The strain used for overproduction of IHF is described (Nash, H.A. *et. al.* *J. Bacteriol.* 169: 4121-4127 (1987)). IHF was purified as described (Rice, P.A. *et. al.* *Cell* 87: 1295-1306 (1996)).

20

**Purification of Native Int.** Native Int was purified from *E. coli* strain HN695 (Lange-Gustafson, B.J. and Nash, H.A. *J. Biol. Chem.* 259:12724-12732 (1984)) by a modification of published procedures (Nash, H.A. *Methods Enz.* 100:210-216 (1983)).

25

**Growth of Cells.** Cells from a glycerol stock of strain HN695 were inoculated into 50 ml of LB broth containing 25  $\mu\text{g}/\text{ml}$  ampicillin in a 250-ml flask. The culture was grown at 31 °C in an air shaker to an  $A_{650}$  of 0.6 to 1.4. This seed culture was used to inoculate six 2.8-L flasks containing 500 ml of growth medium each and cells were grown as just stated. These cultures were used to inoculate 360 L of growth medium in a 500-L fermentor. Cells were

30

grown at 31°C with aeration (190 rpm) and agitation (200 rpm) to an  $A_{650}$  of 0.65, and were harvested in a chilled centrifuge. Cell paste (~ 400 g) was brought to 5 600 ml by addition of ice-cold 50 mM Tris-HCl (pH 7.5) containing 10% (w/v) sucrose and homogenized in a Waring blender at low speed. The slurry was divided into 40-ml aliquots, frozen in dry ice, and stored at -70 °C.

Preparation of Extract. Three tubes of frozen cells (60 g) were thawed at room temperature and placed on ice. To each tube, 2 ml of a 10 mg/ml solution of lysozyme in 250 mM Tris-HCl (pH 7.5) was added, and the tubes were mixed thoroughly. After 35 min on ice, the mixture was centrifuged at 10 32,600  $\times g$  for 45 min. The supernatant was retained (57 ml).

Differential Salt Precipitation. The supernatant was diluted with 50 mM Tris-HCl (pH 7.5) to 100 ml and centrifuged at 4 °C and 41,000 rpm (170,000 x g) for 200 min in a precooled Sorval T865 rotor. The supernatant was decanted, frozen, and stored at -70 °C. The pellet was stored at -70°C. Thawed pellet was resuspended with the aid of a Teflon pestle in Buffer X (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM  $\beta$ -mercaptoethanol, and 10% (w/v) glycerol) + 0.6 M KCl. After adjusting to a volume of 50 ml with the same buffer, the mixture was stirred at 4°C for 1 hr and centrifuged in a Sorval T865 rotor as before. The clear, straw-colored supernatant was carefully removed, frozen in dry ice, and stored at -70°C.

25 Phosphocellulose Chromatography. After thawing, the second supernatant was loaded at 38 cm/hr on a 4.5- ml phosphocellulose column (Whatman P-11) equilibrated in Buffer X + 0.6 M KCl and the column was washed with 5 column volumes of Buffer X + 0.6 M KCl. The column was developed with a 10-column volume linear gradient of Buffer X + 0.6 M KCl to Buffer X + 1.7 M KCl at 19 cm/hr. Int-containing fractions eluting between 0.7 and 1.1 M KCl were pooled and stored at -70 °C.

30 Hydroxyapatite Chromatography. The phosphocellulose pool was loaded at 38 cm/hr on a 1.5-ml hydroxyapatite column (Bio-Rad, ceramic, type II) equilibrated in Buffer X + 0.6 M KCl. The pool was diluted with Buffer X to match the ionic strength of Buffer X + 0.6 M KCl before loading. The column

10 was washed with buffer X + 1 M KCl. Int was eluted at 19 cm/hr with a 10-  
column volume linear gradient of Buffer X + 0.6 M KCl to Buffer X + 0.6 M  
KCl + 0.025 M KPO<sub>4</sub>. Int-containing fractions were pooled, BSA was added to  
2 mg/ml, and the pool was frozen at -70 °C.

5

#### *Purification of Stimulatory Protein as a Side Fraction of a Native Int Preparation*

10 Cells were grown and harvested and cell extract was prepared as described  
in the Materials and Methods section Purification of Native Int. The clarified cell  
extract (~ 60 ml) was diluted to 100 ml with Buffer X (see section: Purification  
of Native Int) and centrifuged at 4 °C at 41,000 rpm in a Sorval T865 rotor for  
200 min. The supernatant was divided into 25 ml aliquots in 50 ml conical tubes  
and submerged into a boiling water bath for 30 minutes. The heated suspension  
was centrifuged at 27,000 x g for 45 minutes. The supernatant was collected and  
diluted with Buffer X + 1.7 M KCl to match the ionic strength of Buffer X +  
0.6 M KCl and loaded at 15 cm/hr onto a 18 ml phosphocellulose (Whatman P-  
11) column (1.6 x 9 cm) which had been equilibrated in Buffer X + 0.6 M KCl.  
The column was washed with 10 column volumes of Buffer X + 0.6 M KCl and  
developed with a 10-column volume linear gradient of Buffer X + 0.6 M KCl to  
Buffer X + 1.7 M KCl. Fractions were stored at -70° C. SDS PAGE analysis of  
aliquots of the fractions revealed a single protein band migrating with an apparent  
molecular weight of 11 kDa. The protein eluted at 1.2 M KCl. Fractions  
containing the 11-KDa protein were pooled and diluted with Buffer X to match  
the ionic strength of Buffer X + 0.2 M KCl. The diluted pool was loaded at 76  
25 cm/hr onto a 1 ml Mono S column (Pharmacia) equilibrated in Buffer X + 0.2 M  
KCl. The protein was eluted with Buffer X + 1.0 M KCl. Fractions containing  
the peak of 11-KDa protein were pooled and stored at -70° C. The protein was  
subjected to amino-terminal amino acid sequence analysis as described in  
Materials and Methods section Amino-Terminal Amino Acid Sequence Analysis  
30 of Stimulatory Proteins and found to be ribosomal protein S20.

**Purification of Stimulatory Proteins from Cells Producing Native Int**

Cells were grown and harvested as described in Materials and Methods section Purification of Native Int. Cell slurry (60 g cells) was thawed at room temperature and placed on ice. A 20 mg/ml solution of lysozyme in 250 mM Tris-HCl (pH 7.4) was added in a volume 1/20 the volume of cells. After 40 minutes on ice with occasional mixing, KCl was added to a final concentration of 0.6 M. The slurry was divided into 25 ml aliquots in 50 ml conical tubes and submerged in a 72°C water bath for 25 minutes. The suspension was spun at 27,000 x g for 45 minutes. The supernatant was loaded at 15 cm/hr onto a 10 ml phosphocellulose column (Whatman P-11) (1.6 x 5 cm) equilibrated in Buffer X + 0.6 M KCl. The column was washed with 10 column volumes of Buffer X + 0.6 M KCl and developed with a 10-column volume linear gradient of Buffer X + 0.6 M KCl to 1.7 M KCl. The fractions were assayed for ability to stimulate  $\lambda$  integrase activity (see Materials and Methods section Integrative Recombination Gel Assay). Two peaks of stimulating activity were found. Two pools were made, from fractions eluting at ~ 0.8 M KCl (Pool 1) and from fractions eluting at ~ 1.2 M KCl (Pool 2), and stored at -70° C.

The pools were processed separately on Mono S. Each pool was diluted with Buffer X to match the ionic strength of Buffer X + 0.2 M KCl and loaded at 76 cm/hr onto a 1 ml Mono S column (Pharmacia) equilibrated with Buffer X + 0.2 M KCl. The column was washed with 10 column volumes of Buffer X + 0.2 M KCl and developed with a 20-column volume linear gradient of Buffer X + 0.2 M KCl to Buffer X + 1.7 M KCl. Fractions were stored at -70° C.

The fractions from each column were assayed for ability to stimulate  $\lambda$  integrase activity. Pool 1 from phosphocellulose was fractionated into two activity peaks by Mono S. The primary protein band in the first peak (Figure 18, lanes A and B) was determined by N-terminal amino acid sequence analysis to be ribosomal protein L27 (see Materials and Methods section Amino-Terminal Amino Acid Sequence Analysis of Stimulatory Proteins). The second peak eluting later in the gradient was found to be composed of two major protein bands by SDS PAGE analysis (Figure 18, lanes C and D). One protein co-migrated with

L27 and the other migrated more slowly than L27 and S20 (lane E). Pool 2 from phosphocellulose was fractionated into one peak of activity by Mono S which eluted at a slightly higher salt concentration than the second peak of Pool 1 on Mono S. The main protein in this activity peak co-migrated during SDS-PAGE analysis with S20 protein (Figure 18, lanes F and G).

5

#### *Amino-Terminal Amino Acid Sequence Analysis of Stimulatory Proteins*

10 Protein samples were subjected to SDS PAGE as described in Materials and Methods section SDS PAGE. The gel was equilibrated in transfer buffer (0.05 M Tris, 0.04 M boric acid, 0.5 mM EDTA, 20% (v/v) methanol (pH 8.4)). PVDF membrane (Immobilon P from Millipore) was prepared according to manufacturer's instructions and equilibrated in transfer buffer. The protein was transferred to the membrane using a BioRad mini blotting apparatus at 100 volts for 1 hour. The membrane was stained with Coomassie Blue R-250 staining solution and destained in 100% (v/v) methanol. The membrane was air dried and the stained protein band was excised from the membrane and stored in a 1.5-ml microcentrifuge tube.

20 Amino-terminal amino acid sequence analysis was performed on membrane bound protein samples by automated Edman sequence analysis by the HHMI Biopolymer Laboratory, W.M. Keck Foundation, New Haven, CT.

25

#### *Cloning of Int-His<sub>6</sub>*

The following two oligonucleotides were used to clone the Int gene: TAT TAT TAT **CAT ATG GGA CGA CGT CGA AGT CAT GAG CGC CGG GAT** (SEQ ID NO:1) and A TTA TTA **AGC TTA ATG GTG ATG ATG GTG ATG TTT GAT TTC AAT TTT GTC CCA CTC** (SEQ ID NO:2). The oligonucleotides were used to generate a 1,092-bp PCR amplification product using  $\lambda$  DNA as the template. DNA was amplified (Materials and Methods section PCR) during 8 cycles composed of the following

30

5 steps: 95 °C for 15 seconds, 55°C for 15 seconds, and 72 °C for 90 seconds. The 1,092-bp PCR product was digested with *Nde*I and *Hind*III and cloned into the *Nde*I and *Hind*III sites of plasmid pTRCN2 (Figure 10) in an *E. coli* DH10B host. This construct is called pTRCN2INT2 (Figure 11). The Int gene is under control of a pTRC promoter and contains a sequence coding for a His<sub>6</sub> tag at the carboxy end of the protein. The DNA sequence of the Int gene in pTRCN2INT2 was determined and found to match the published sequence, except as modified below. Arg codons AGA and AGG originally coding for Arg at positions 3 and 4 were changed to CGA and CGT, respectively, which are Arg codons more frequently used in *E. coli*.

10

#### ***Purification of Int-His<sub>6</sub>***

Int-His<sub>6</sub> was purified from *E. coli* DH10B cells bearing plasmid pTRCN2INT2 (see Materials and Methods section Cloning of Int-His<sub>6</sub>).

15 *Growth of Cells.* To prepare seed stocks, *E. coli* DH10B cells bearing plasmid pTRCN2INT2 were grown at 30 °C in Buffered Rich medium + 100 µg/ml ampicillin to an  $A_{590}$  ~2. Culture was mixed 1:1 with 50% glycerol. The mixture was aliquoted by 1 ml into cryovials on ice and then stored at -80 °C.

20 For a small scale growth, cells from a frozen glycerol stock were inoculated into 2 x 50 ml Buffered Rich medium + 100 µg/ml ampicillin in 2 x 250-ml bottom-baffled shake flasks. Cells were grown for 16.5 hours at 30 °C and 250 rpm to an  $A_{590}$  of ~4.0. Twenty-five ml of the primary shake flask growth was used to inoculate each of 4, 2.8-L bottom-baffled Fernbach flasks containing 1 L of Buffered Rich medium + 100 µg/ml ampicillin (for an initial  $A_{590}$  of ~0.1). Cultures were grown at 30 °C until an  $A_{590} = 1.0$  to 1.5 was achieved. The 25 cultures were induced by adding IPTG to 1 mM. Growth was continued for 2 hr at 30 °C. The culture was chilled by icing in 4 x 1 L centrifuge bottles and harvested by centrifugation at 4,500 rpm (5,895 x g) and 4 °C for 12 minutes. Each pellet was washed by resuspension in ~7 ml 50 mM Tris-HCl (pH 8.0), 100 mM NaCl at 4 °C and re-spun. The pellets were frozen and stored at -80 °C.

30

For a large scale growth, 50 ml of Buffered Rich medium + 100  $\mu$ g/ml ampicillin in a 250 ml bottom baffled shake flask was inoculated with 1 ml of a frozen seed. Cells were grown at 30 °C and 250 rpm to an  $A_{590}$  of 0.8 to 1.2. The entire 50 ml was inoculated into 500 ml Buffered Rich medium + 100  $\mu$ g/ml in a 2.8-L bottom-baffled Fernbach. Growth was continued at 30 °C and 250 rpm to an  $A_{590}$  = 0.8 to 1.2.

10 L of Buffered Rich medium + 100  $\mu$ g/ml ampicillin in a 14-L vessel was inoculated with all 500 ml of culture. Temperature was maintained at 30 °C. Dissolved oxygen levels were controlled at >30% and pH at 7 +/- 0.3. At  $A_{590}$  = 1.5 to 2.0 the culture was induced by adding IPTG to 1 mM. Growth was continued for 2 hr at 30 °C. The vessel was chilled and harvested by centrifugation in a Sharples centrifuge. Cell paste was frozen and stored at -80C.

*Purification.* Frozen cells (20 g) were thawed on ice and suspended in 40 ml of Tris-HCl (pH 8.0)-10% (w/v) sucrose. Cells were disrupted on ice by sonication (4, 30 second bursts at 70% maximum setting), and the extract was centrifuged at 27,000 x g for 30 minutes at 4 °C. The supernatant was collected. The supernatant was mixed with 20 ml (packed volume) of Chelating Sepharose (Pharmacia) charged with  $NiSO_4$  and equilibrated with Buffer A (50 mM Tris-HCl (pH 8.0), 0.3 M NaCl, 10% (v/v) glycerol). The slurry was transferred to 50-ml conical tubes and was gently rocked for 30 minutes at 4 °C. The slurry was then packed into a 1.6 cm column and attached to an FPLC system (Pharmacia). The column was washed with 20 column volumes of Buffer A + 20 mM Imidazol at 30 cm/hr. The protein was eluted with a 15-column volume linear gradient from Buffer A + 20 mM Imidazol to Buffer A + 500 mM Imidazol. Fractions were analyzed by SDS PAGE. Fractions containing Int-His<sub>6</sub> were pooled and 0.5 M EDTA was added to a final concentration of 1 mM. The pool was then transferred to 10,000 molecular weight cut off (MWCO) dialysis tubing and dialyzed against 50 volumes of Buffer B (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 10% (v/v) glycerol, and 1 mM  $\beta$ -mercaptoethanol). The dialyzed pool was loaded at 38 cm/hr onto a 2 ml (1 x 1 cm) EMD-SO<sub>4</sub> (EM Separations) column equilibrated in Buffer B + 0.2 M NaCl. The column was washed with 10

5

10

20

25

30

column volumes of Buffer B + 0.2 M NaCl at 76 cm/hr and developed with a 15-  
column volume linear gradient from Buffer B + 0.2 M NaCl to Buffer B + 1.6 M  
NaCl. Int-His<sub>6</sub> eluted at approximately 1.1 M NaCl based upon analysis by SDS  
PAGE. The peak fractions were pooled and the pool was transferred to 10,000  
MWCO dialysis tubing and dialyzed against 100 volumes of Buffer C (Buffer B  
minus EDTA). The dialyzed pool was loaded at 38 cm/hr onto a 1 ml (0.5 x 1 cm)  
hydroxyapatite column (Type II, BioRad) equilibrated in Buffer C. The column  
was washed with 10 column volumes of Buffer C + 1 M NaCl and developed  
with 10 column volumes of Buffer C + 0.6 M NaCl + 25 mM KPO<sub>4</sub> at 19 cm/hr.  
The fractions were analyzed by SDS PAGE and the peak fractions containing Int-  
His<sub>6</sub> were pooled. The pool was transferred to 10,000 MWCO dialysis tubing and  
was dialyzed against 200 volumes of 50 mM Tris-HCl (pH 7.5), 50 mM NaCl,  
0.05 mM EDTA, 50% (v/v) glycerol, and 1 mM DTT overnight at 4° C. The  
final sample was stored at -70° C.

#### *Cloning of Xis-His<sub>6</sub>*

The following two oligonucleotides were used to clone the Xis gene: TAT  
TAT TAT **CAT ATG** TAC TTG ACA CTT CAG GAG (SEQ ID  
NO:3) and ATT ATT **AAG CTT** ATT AAT GGT GAT GAT GGT  
GAT GTG ACT TCG CCT TCT TCC CAT T (SEQ ID NO:4). The  
oligonucleotides were used to generate a 219-bp PCR product using λ DNA as  
the template. DNA was amplified (Materials and Methods section PCR) during  
15 cycles composed of the following steps: 95 °C for 15 seconds, 55 °C for 15  
seconds, and 72 °C for 60 seconds. The 219-bp PCR product was digested with  
*Nde*I and *Hind*III and cloned into the *Nde*I and *Hind*III site of pTRCN2 (Figure  
10). The resulting construct was called pTRCN2XIS1 (Figure 12). The Xis gene  
is under control of a pTRC promoter and contains a sequence coding for a His<sub>6</sub>  
tag at the carboxy end of the protein. The DNA sequence of the Xis gene in  
pTRCN2XIS1 was determined and found to match the published sequence.

*Purification of Xis-His<sub>6</sub>*

Xis-His<sub>6</sub> was purified from *E. coli* Stbl 2 cells bearing plasmid pTRCN2XIS1 (see Materials and Methods section Cloning of Xis-His<sub>6</sub>).

5 *Growth of Cells.* To prepare seed stocks, *E. coli* Stbl 2 cells bearing plasmid pTRCN2XIS1 were grown at 37 °C in Buffered Rich medium + 100 µg/ml ampicillin to an  $A_{590}$  ~3. Culture was mixed 1:1 with 50% glycerol. The mixture was aliquoted by 1 ml into cryovials on ice and then stored at -70 °C.

10 For small scale growths, cells from a frozen glycerol stock were inoculated into 50 ml Buffered Rich medium + 100 µg/ml ampicillin in a 250-ml bottom-baffled shake flask. Cells were grown for 17 hours at 37 °C and 250 rpm to an  $A_{590}$  of ~4.0.

15 12 ml of the primary shake flask growth was used to inoculate each of 4, 2.8-L bottom-baffled Fernbach flasks containing 1 L of Buffered Rich medium + 100 µg/ml ampicillin (for an initial  $A_{590}$  of ~0.05). Cultures were grown at 37 °C until an  $A_{590}$  = 1.5 to 2.0 was achieved. The cultures were induced by adding IPTG to 1 mM. Growth was continued for 2 hr at 37 °C. The culture was chilled by icing in 4 x 1 L centrifuge bottles and harvested by centrifugation at 4,500 rpm (5,895xg) and 4 °C for 15 minutes. Each pellet was washed by resuspension in ~20 ml used medium and re-spun. The pellets were frozen and stored at -70 °C.

20 For a large scale growth, a 50 ml culture of Buffered Rich medium + 100 µg/ml ampicillin in a 250-ml bottom baffled shake flask was inoculated with 1 ml of a frozen seed. Cells were grown at 37 °C and 250 rpm to an  $A_{590}$  of 0.6 to 1.4. The entire 50 ml was inoculated into 500 ml Buffered Rich medium + 100 µg/ml ampicillin in a 2.8-L bottom-baffled Fernbach. Growth was continued at 37 °C and 250 rpm to an  $A_{590}$  = 0.6 to 1.4. Ten L of Buffered Rich medium + 100 µg/ml ampicillin in a 14-L vessel was inoculated with all 500 ml of culture. Temperature was maintained at 37 °C. Dissolved oxygen levels were controlled at >30% and pH at 7 +/- 0.3. At  $A_{590}$  = 1.5 to 2.0 the culture was induced by adding IPTG to 1 mM. Growth was continued for 2 hr at 37 °C. The vessel was chilled and harvested by centrifugation in a Sharples centrifuge. Cell paste was frozen and stored at -70 °C.

*Purification.* Frozen cells (20 g) were thawed on ice and suspended in 20 ml of 50 mM Tris-HCl (pH 8.0), 10% (w/v) sucrose, 0.002 mg/ml leupeptin, 0.002 mg/ml pepstatin A, 0.8 mg/ml benzamide, and 0.05 mg/ml Pefablock. Cells were disrupted by sonication (5 second bursts at 80% of the maximum setting alternated with 5 seconds off for 3 minutes). The extract was centrifuged at 27,000 X g for 30 minutes at 4° C and the supernatant was collected. The supernatant was loaded at 30 cm/hr onto a 20-ml column (1.6 x 10 cm) of Chelating Sepharose (Pharmacia) charged with NiSO<sub>4</sub> and equilibrated with Buffer D (50 mM Tris-HCl (pH 7.5), 0.4 M NaCl, and 10 % (v/v) glycerol) + 5 mM Imidazol. The column was washed with 20 column volumes of Buffer D + 5 mM Imidazol at 30 cm/hr and developed with a 15-column volume linear gradient from Buffer D + 5 mM Imidazol to Buffer D + 450 mM Imidazol at 12 cm/hr. Fractions were analyzed by SDS PAGE. Peak fractions containing the Xis-His<sub>6</sub> protein were pooled and 0.5 M EDTA and 1 M DTT were added to final concentrations of 1 mM and 4 mM, respectively. The pool was then loaded at 38 cm/hr onto a 5.5 ml (1.0 x 7.0 cm) EMD-SO<sub>4</sub> (EM Separations) column equilibrated in Buffer E (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 10% (v/v) glycerol, and 4 mM DTT) + 0.4 M NaCl. The column was washed with 10 column volumes of Buffer E + 0.4 M NaCl at 76 cm/hr and developed with a 10-column volume linear gradient from Buffer E + 0.4 M NaCl to Buffer E + 2 M NaCl at 15 cm/hr. Fractions were analyzed by SDS PAGE. Xis-His<sub>6</sub> elutes in a broad peak at approximately 1.1- 1.8 M NaCl. The peak fractions containing Xis-His<sub>6</sub> were pooled. The pool was diluted with Buffer E to match the ionic strength of Buffer E + 0.2 M NaCl and loaded at 152 cm/hr onto a 1 ml (0.5 x 5.0 cm) Mono S (Pharmacia) column equilibrated in Buffer E + 0.2 M NaCl. The column was washed with 10 column volumes of Buffer E + 0.2 M NaCl. Xis-His<sub>6</sub> was eluted with 10 column volumes of Buffer E + 2.0 M NaCl at 61 cm/hr. Fractions were analyzed by SDS PAGE and the peak fractions containing Xis-His<sub>6</sub> were pooled. The pool was transferred to a 2,000 molecular weight cut off dialysis cassette (Pierce) and was dialyzed against 200 volumes of 50 mM Tris-

HCl (pH 7.5), 50 mM NaCl, 0.05 mM EDTA, 50% (v/v) glycerol, and 1 mM DTT overnight at 4° C. The final sample was stored at -70° C.

**Cloning of S20**

5 The following two oligonucleotides were used to clone the S20 gene:  
TAT TAT TAT **CAT ATG** GCT AAT ATC AAA TCA GCT AAG  
(SEQ ID NO:5) and ATT ATT **GGA TCC** ATT AAG CCA GTT  
TGT TGA TCT (SEQ ID NO:6). The oligonucleotides were used to generate  
10 a 267-bp PCR product using *E. coli* chromosomal DNA as template. DNA was  
amplified (Materials and Methods section PCR) during 15 cycles composed of  
the following steps: 95 °C for 15 seconds, 50 °C for 15 seconds, and 67 °C for 30  
seconds. The 267-bp PCR product was digested with *Nde*I and *Bam*H I and  
cloned into the *Nde*I and *Bam*H I sites of pTRCN2 (Figure 10) in *E. coli* DH10B.  
The resulting construct was called pTRCN2S20AA (Figure 13). The S20 gene is  
under control of a pTRC promoter. The DNA sequence of the S20 gene in  
pTRCN2S20AA was determined and found to match the published sequence,  
except as noted below. The initiation codon was changed from TTG to ATG  
during cloning to enhance expression. pTRCN2S20AA was digested with *Nde*I  
and *Bam*H I to generate a 267-bp fragment that was cloned into the *Nde*I and  
15 *Bam*H I sites of pET12A ( Novagen) in *E. coli* strain BL21DE3. The resulting  
construct was called pET12AS20AA (Figure 14). The S20 gene is under control  
of a T7 promoter.

**Purification of Recombinant S20**

25 S20 was purified from *E. coli* BL21DE3 bearing plasmid pET12AS20AA  
(see Materials and Methods section Cloning of S20).

*Growth of Cells.* Cells from a glycerol stock of BL21DE3 bearing plasmid  
pET12AS20AA were inoculated into 3 ml of LB broth containing 100 µg/ml  
ampicillin. This inoculum was diluted into LB broth + 100 µg/ml ampicillin  
30 1:100 and the 300-ml culture was grown overnight at 30 °C. The  $A_{650}$  of the  
culture should not exceed 1.0. This culture was used to innoculate 10 flasks

*sub C2*

containing 500 ml each of Circlegrow (BIO 101) plus 100  $\mu$ g/ml ampicillin plus 1 mM MgSO<sub>4</sub>. Cells were grown at 37 °C until the  $A_{650}$  was 0.5 and expression of S20 was induced by the addition of IPTG to 0.5 mM. After growth at 37 °C for 4 hours, cells were harvested by centrifugation at 4 °C and stored at -70 °C.

5 *Purification.* Frozen cells (10 g) were thawed on ice and suspended in

25 25 ml of 50 mM Tris-HCl (pH 7.5), 0.2 mM EDTA, 10% (v/v) glycerol, 0.2 mM DTT, 0.2  $\mu$ g/ml leupeptin, and 1 mM PMSF. Cells were then disrupted by sonication (5 second bursts at 80% of the maximum setting alternated with 5 seconds off for 1.5 minutes). NaCl (5.0 M) was then added to a final concentration of 0.67 M. The slurry was mixed by inverting the container and then placed on ice for 10 minutes. The mixture was centrifuged at 27,000 X g for 30 minutes at 4 °C and the supernatant was collected. The supernatant was diluted with Buffer B (50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 10% (v/v) glycerol, 1 mM  $\beta$ -mercaptoethanol) to match the ionic strength of Buffer B + 0.3 M NaCl and then loaded at 30 cm/hr onto a 7.5 ml (1.8 x 3.7 cm) EMD-SO<sub>4</sub> (EM Separations) column equilibrated in Buffer B + 0.3 M NaCl. The column was washed with 10 column volumes of Buffer B + 0.3 M NaCl at 30 cm/hr and developed with a 15-column volume linear gradient from Buffer E + 0.3 M NaCl to Buffer E + 1.8 M NaCl at 30 cm/hr. Fractions were analyzed by SDS PAGE. S20 eluted at approximately 0.9 M NaCl. The fractions containing the peak of S20 were pooled. The pool was transferred to a 2,000 molecular weight cut off dialysis cassette (Pierce) and dialyzed against 200 volumes of 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 0.05 mM EDTA, 50% (v/v) glycerol, and 1 mM DTT overnight at 4 °C. The final sample was stored at -70 °C.

25 **Integrative Recombination Gel Assay**

30 Reaction mixtures (10  $\mu$ l final volume) for monitoring integrative recombination (defined as containing linearized *attB* and supercoiled *attP* DNA substrates) by agarose gel electrophoresis were incubated at 25 °C for 45 minutes. Reactions were initiated by adding 1  $\mu$ l of Int or Int-His<sub>6</sub> (contained in 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 600 mM KCl, 2 mg/ml BSA, and 10% (v/v)

glycerol) plus or minus potential stimulatory proteins to a mixture containing 20 mM Tris-HCl (pH 8.0), 5 mM spermidine, 50 µg/ml BSA, 125 ng linearized pBB105, 125 ng supercoiled pHN894, and 12.5 ng IHF. Incubation was stopped by raising the temperature to 70 °C for 10 minutes and then adding 2.5 µl of 25%(w/v) Ficoll 400, 0.5% (w/v) SDS, and 0.00625% (w/v) bromophenol blue. In some cases, reaction mixtures were treated with proteinase K (10 to 20 µg at 25 °C for 15 minutes). Samples were analyzed by electrophoresis in a 1% agarose minigel cast in 40 mM Tris-acetate (pH 8.3)- 1 mM EDTA (TAE) and 1 µg/ml ethidium bromide and run in TAE at 105 V for 30 minutes. Recombination activity is indicated by the appearance of a DNA band migrating at 10,201 bp. A unit of Int activity was defined as described (Nash, H.A. *Methods Enz.* 100: 210-216 (1983)).

#### *Excise Recombination Gel Assay*

Reaction mixtures (10 µl final volume) for monitoring excise recombination (defined as containing linearized *attL* and supercoiled *attR* DNA substrates) by agarose gel electrophoresis were incubated at 25 °C for 45 minutes. Reactions were initiated by adding 1 µl of Int or Int-His<sub>6</sub> (contained in 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 600 mM KCl, 2 mg/ml BSA, and 10% (v/v) glycerol) plus or minus potential stimulatory proteins to a mixture containing 20 mM Tris-HCl (pH 8.0), 5 mM spermidine, 50 µg/ml BSA, 125 ng linearized pHN872, 125 ng supercoiled pHN868, 12.5 ng IHF, and 28 ng Xis or Xis-His<sub>6</sub>. Incubation was stopped by raising the temperature to 70 °C for 10 minutes and then adding 2.5 µl of 25%(w/v) Ficoll 400, 0.5% (w/v) SDS, and 0.00625% (w/v) bromophenol blue. In some cases, reaction mixtures were treated with proteinase K (10 to 20 µg at 25 °C for 15 minutes). Samples were analyzed by electrophoresis in a 1% agarose minigel cast in 40 mM Tris-acetate (pH 8.3)- 1 mM EDTA (TAE) and 1 µg/ml ethidium bromide and run in TAE at 105 V for 30 minutes. Recombination activity is indicated by the appearance of a DNA band migrating at 9,991 bp.

#### **Integrative Recombination Colony-Forming Assay**

Reaction mixtures (20  $\mu$ l final volume) for monitoring integrative recombination (defined as containing linearized *attB* and supercoiled *attP* DNA substrates) by transformation of *E. coli* were incubated at 25 °C for 45 minutes. Reactions were initiated by adding 4  $\mu$ l of Int or Int-His<sub>6</sub> (contained in 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA, 200  $\mu$ g/ml BSA, and 50% (v/v) glycerol) plus or minus S20 to a mixture containing 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 2.5 mM spermidine, 0.25 mM EDTA, 200  $\mu$ g/ml BSA, 100 ng linearized pEZC7501, 100 ng supercoiled pEZ13835, and 10 ng IHF. Incubation was stopped by raising the temperature to 70 °C for 10 minutes. Proteinase K (4  $\mu$ g in 1  $\mu$ l) was added and after 10 minutes at 37 °C the mixture was centrifuged (14,000 rpm for 30 seconds). The mixture (1  $\mu$ l) was used to transform 100  $\mu$ l of ME DH5 $\alpha$  *E. coli* competent cells (LTI) in a sterile polypropylene tube on ice. After 30 minutes on ice, the tube was heat shocked in a 42 °C water bath for 45 seconds. The tube was then placed on ice for 2 minutes. S.O.C. medium (0.9 ml) was added to the tube, and the tube was placed in a shaker for 60 minutes at 37 °C and 225 rpm. Aliquots (10 and 100  $\mu$ l) of the transformed cells were spread on separate agar plates prepared in LB medium + 100  $\mu$ g/ml kanamycin, and the plates were incubated at 37 °C for 16 to 24 hours. Kanamycin-resistant colonies arise only as the result of an integrative recombination event.

#### **Excisive Recombination Colony-Forming Assay**

Reaction mixtures (20  $\mu$ l final volume) for monitoring excisive recombination (defined as containing linearized *attR* and supercoiled *attL* DNA substrates) by transformation of *E. coli* were incubated at 25 °C for 45 minutes. Reactions were initiated by adding 4  $\mu$ l of Int or Int-His<sub>6</sub> (contained in 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA, 200  $\mu$ g/ml BSA, and 50% (v/v) glycerol) plus or minus S20 to a mixture containing 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 2.5 mM spermidine, 0.25 mM EDTA, 200  $\mu$ g/ml BSA, 100 ng linearized pEZC8402, 100 ng supercoiled pEZ11104, 12.5 ng IHF, and 28 ng Xis

or Xis-His<sub>6</sub>. Incubation was stopped by raising the temperature to 70 °C for 10 minutes. Proteinase K (4 µg in 1 µl) was added and after 10 minutes at 37 °C the mixture was centrifuged (14,000 rpm for 30 seconds). A portion of the reaction mixture (10.5 µl) was diluted with 89.5 µl of T<sub>10</sub>E<sub>1</sub>. The diluted mixture (1 µl) was used to transform 100 µl of ME DH5 $\alpha$  *E. coli* competent cells (LTI) in a sterile polypropylene tube on ice. After 30 minutes on ice, the tube was heat shocked in a 42 °C water bath for 45 seconds. The tube was then placed on ice for 2 minutes. S.O.C. medium (0.9 ml) was added to the tube, and the tube was placed in a shaker for 60 minutes at 37 °C and 225 rpm. Aliquots (10 and 100 µl) of the transformed cells were spread on separate agar plates prepared in LB medium + 100 µg/ml ampicillin, and the plates were incubated at 37 °C for 16 to 24 hours. Ampicillin-resistant colonies arise only as the result of an excisive recombination event.

## RESULTS

### PART I: *Restoration of Integrase Activity by Mixing with Cell Extract Components*

*Restoration of Int Activity by Column Fractions.* Purification of Int overexpressed in *E. coli* involved differential salt precipitation followed by phosphocellulose and hydroxyapatite chromatography (Materials and Methods). When we attempted to purify native Int by this procedure, we found that Int integrative recombination activity (determined as described in Materials and Methods section, Integrative Recombination Gel Assay) was maintained through the phosphocellulose chromatography step, but was lost during the final hydroxyapatite chromatography step. No activity was found in any hydroxyapatite column fraction. This was not caused by loss of Int protein during chromatography, since SDS-PAGE analysis of the hydroxyapatite fractions revealed the presence of a single protein of molecular weight 40 KDa, consistent with the bound protein being Int. Fractions containing the peak of the 40-KDa

protein were pooled and the pool was assayed for integrative recombination activity. As the results shown in Table 2 indicate, no activity was observed.

TABLE 2: SUMMARY OF PURIFICATION OF NATIVE Int

Purification Step	Total Units	Total Protein (mg)	Specific Activity (U/mg)
Crude Extract	228,000	1,294	176
Differential salt precipitation	67,000	153	441
Phosphocellulose	21,000	6.7	3,134
Hydroxyapatite	0	0.2	--
Hydroxyapatite + stimulatory protein(s)	~30,000	0.2	~150,000

Examination of the proteins in the phosphocellulose pool by SDS PAGE revealed the presence of Int (40 KDa) and a number of smaller proteins (at least six) in the 5 to 17 KDa range. *E. coli* DNA binding proteins that stimulate Int activity, such as HU, fall in this small size range (Segall, A.M. et. al, *EMBO J.* 13: 4536-4548 (1994)). We therefore hypothesized that this preparation of Int required additional component(s) for activity beyond the IHF already present in recombination reaction mixtures (Materials and Methods). Further, the chromatography results suggested that this component(s) coeluted with Int from phosphocellulose, but was not bound by hydroxyapatite. To test this hypothesis, the material from the original phosphocellulose pool that did not bind to hydroxyapatite was fractionated again on a phosphocellulose column. Samples from fractions from this column were assayed for ability to restore integrative recombination activity to the inactive Int pooled from the hydroxyapatite column. We found that fractions eluting from the phosphocellulose column at around

1.0 M KCl contained a component(s) that restored recombination activity to the inactive Int (Figure 15). The fractions with the greatest stimulatory activity (Fraction Numbers 15 through 18 in Figure 15) were used for further characterization. Unit assay of the Int hydroxyapatite pool in the integrative recombination assay in the presence of an optimal amount of this stimulatory material indicated that greater than 100% of the Int activity present in the phosphocellulose pool was present in the hydroxyapatite pool when the stimulatory component(s) was present in the unit assay (Table 2).

Characterization of the Stimulatory Component(s). SDS PAGE analysis of the stimulatory fractions from the second phosphocellulose column showed multiple small protein bands, two of which appeared similar in size to the subunits of authentic IHF (Figure 15). On the chance that the concentration of IHF being used in the integrative recombination gel assay was not optimal, a careful titration of IHF was carried out with inactive Int in the presence and absence of stimulatory material from the phosphocellulose column. We found that no amount of IHF alone, from 12.5 to 1,250 ng, stimulated inactive Int. In contrast, the combination of IHF at 12.5 ng and the component(s) from the phosphocellulose column did restore Int activity.

Treatment of the stimulatory component(s) with DNase I or RNase A did not diminish ability to stimulate Int. Placing the component(s) in a boiling water bath for 30 minutes also had no effect. However, treatment with proteinase K eliminated ability to stimulate, indicating the stimulatory component(s) was protein that could withstand high temperature.

## PART II: Purification and Identification of the Stimulatory Proteins

Purification from a Side Fraction. We wished to identify the protein(s) in extracts of *E. coli* expressing native Int that stimulate its recombinase activity. Purification was monitored by detecting the presence of Int stimulatory protein using the integrative recombination gel assay (Materials and Methods) and inactive Int, purified as just described (Materials and Methods section

Purification of Native Int and Results section PART I: Restoration of Integrase Activity by Mixing with Cell Extract Components). We took advantage of the fact that extracts could be heated to boiling water temperatures without affecting adversely the stimulatory activity. Heating served several purposes. First, any active Int present during early purification steps would be irreversibly inactivated, eliminating interference in the gel recombination assay. Second, many *E. coli* proteins in crude extracts precipitate at high temperature; thus heating facilitates purification of those proteins that remain soluble.

The side fractions generated early in the native Int purification (Materials and Methods section Purification of Native Int) were heated to 100 °C, clarified by centrifugation, and assayed for ability to stimulate inactive Int. The supernatant from the first high speed centrifugation in the differential salt precipitation step was found to have the most stimulatory activity. Using this supernatant as starting material, a stimulatory protein was purified as described in Materials and Methods section Purification of Stimulatory Protein as a Side Fraction of a Native Int Preparation. A near homogeneous 11-KDa protein was purified after two column chromatography steps (Figure 16) that stimulated inactive Int in the gel recombination assay (Figure 17).

The 11-KDa protein was sent to the HHMI Biopolymer Laboratory, W.M. Keck Foundation, for amino terminal amino acid sequence analysis (Materials and Methods section Amino-Terminal Amino Acid Sequence Analysis of Stimulatory Proteins). The sequence was found to be Ala-Asn-Ile-Lys-Ser-Ala-Lys-Lys-Arg-Ala-Ile-Gln-Ser-Glu (SEQ ID NO:7). Search of the GenBank sequence data base revealed that this sequence matches amino acids 2 through 15 of *E. coli* 30S ribosomal protein S20 (Mackie, G.A. *J. Biol. Chem.* 256:8177-8182 (1981)). S20 is a very basic protein of 86 amino acids. In *E. coli*, S20 appears to be involved in association of the 30S ribosomal subunit with the 50S subunit and in formation of the 30S subunit translation initiation complex with fMet-tRNA and mRNA (Gotz, F. et. al *Biochim. Biophys. Acta* 1050: 93-97 (1990)). The gene for S20 was cloned, overexpressed, and purified (see Materials

and Methods sections Cloning of S20 and Purification of Recombinant S20). The ability of recombinant S20 to stimulate Int was tested (see Results, PART III).

*Purification from Total Cell Extract.* Since we were able to identify one small, heat resistant, nucleic acid binding protein in extracts of *E. coli* that stimulates Int activity, we asked if there were others. Using the gel recombination assay with inactive Int to assay for stimulation of Int, and starting with total *E. coli* cell extract, purification of stimulatory activity was repeated (see Materials and Methods section Purification of Stimulatory Proteins from Cells Producing Native Int). Again, phosphocellulose followed by Mono S chromatography was used to fractionate heated *E. coli* extract. A second stimulatory protein was identified that migrated on SDS PAGE slightly faster than S20 (Figure 18). This protein was also sent to the HHMI Biopolymer Laboratory, W.M. Keck Foundation, for sequence analysis. The sequence was found to be Ala-His-Lys-Lys-Ala-Gly-Gly-Ser-Thr-Arg-Asn (SEQ ID NO:8). Search of the GenBank sequence data base revealed that this sequence matches amino acids 2 through 12 of *E. coli* 50S ribosomal protein L27 (Jeong, J.H. et. al, *DNA Seq.* 4: 59-67 (1993)). L27 is a very basic protein of 85 amino acids. The proteins in fraction 18 (lanes A and B of Figure 18), the primary constituent of which was L27, were tested for ability to stimulate Int in the integrative recombination gel assay. Figure 19 shows that these proteins stimulated Int in the recombination assay. However, 10 times more L27 than S20 was required to produce a discernible recombinant DNA product.

### **PART III: Cloning of S20 and Demonstration of Activity**

*Cloning, Overexpression, and Purification of rS20.* We cloned the gene for S20 from *E. coli* DNA under control of a T7 promoter using PCR (see Materials and Methods section Cloning of S20). The recombinant S20 was highly overexpressed and easily purified by EMD-SO<sub>4</sub> chromatography (see Materials and Methods section Purification of Recombinant S20). Approximately

110 mg of near homogeneous recombinant S20 (Figure 20) was purified from 9 g of *E. coli*.

*Characterization of rsS20.* Recombinant S20 stimulated integrative and excisive  $\lambda$  recombination catalyzed by native Int as determined by gel assay (Figure 19), and recombinant S20 also stimulated both integrative and excisive  $\lambda$  recombination catalyzed by recombinant Int-His<sub>6</sub> as determined both by gel assay (Figure 21) and colony-forming assay (Tables 3 and 4). These results confirmed those obtained with native S20; that is, recombinant S20 stimulates the recombinase activity of Int.

TABLE 3: STIMULATION OF INT-HIS<sub>6</sub> BY RECOMBINANT S20 IN AN INTEGRATIVE RECOMBINATION COLONY-FORMING ASSAY\*

Amt. of Recombinant S20 (ng)	Number of Colonies Formed
0	35
313	82
625	255
1,250	233
2,500	5

\*See Materials and Methods for details of assay. All reaction mixtures contained 176 ng Int-His, and 10 ng IHF.

5 TABLE 4: STIMULATION OF INT-HIS<sub>6</sub> BY RECOMBINANT S20 IN  
AN EXCISIVE RECOMBINATION COLONY-FORMING  
ASSAY\*

6 Amt. of Recombinant S20 (ng)	Number of Colonies Formed
0	9
158	86
313	1,392
625	83
1,250	23

10 \*See Materials and Methods for details of assay. All reaction mixtures contained 176 ng Int-His<sub>6</sub>, 12.5 ng IHF, and 28 ng Xis-His<sub>6</sub>.

15 The order of addition of S20 and Int to a reaction appears to be important. Int should be mixed with S20 and the proteins added as a mixture to IHF and DNAs to obtain greatest stimulation of integrative recombination. If S20 is added before Int, or if Int is added before S20, less stimulation is observed. These results suggest S20 might be binding to Int and producing some kind of physical change that enhances its recombinase activity. Gel shift assays show that S20 binds to the DNA substrates in recombination assays. Thus, treatment of recombination assay mixtures containing large amounts of S20 with proteinase K is necessary to avoid trapping of DNA in wells during agarose gel electrophoresis. Titration of the amount of S20 versus number of recombinants obtained in both the integrative (Table 3) and excisive (Table 4) colony-forming recombination assay demonstrated rather sharp optima for amount of S20, particularly in the excisive reaction. The molar ratios of S20 to DNA nucleotides at the optimal amounts of S20 in these assays were 5 to 10 nucleotides per S20 molecule in the integrative reaction and 25 nucleotides per S20 molecule in the excisive reaction. We speculate that the binding footprint for a protein of the size of S20 (~10 kDa) functioning as a monomer is in the range of 5 to 10 nucleotides per molecule of protein. The optimum for the integrative reaction falls in this range, suggesting that for optimal stimulation of the integrative recombination

5 sufficient S20 must be present to coat the DNA. Making the same assumptions, it would appear that in the excisive reaction, the presence of sufficient S20 to coat the DNA inhibits the reaction. In any case, binding of S20 to DNA is probably also exerting an effect on the efficiency of the recombination reaction, just as does the binding of other small nonspecific DNA binding proteins (Segall, A.M. *et. al, EMBO J. 13: 4536-4548 (1994)*).

10 **PART IV. Integrative Recombination Activity of Int and Int-His<sub>6</sub>**

15 We have completed three purifications of native λ Int following a modification (Materials and Methods) of the published purification procedure (Nash, H.A. *Methods Enz. 100: 210-216 (1983)*), and a much larger number of purifications of cloned Int-His<sub>6</sub> by a simpler procedure (Materials and Methods). As a result of characterization of the integrative recombinase activity of these preparations using the gel assay (see Materials and Methods section Integrative Recombination Gel Assay), we can draw several general conclusions about the activity of Int in the presence and absence of S20. First, preparations of Int or Int-His<sub>6</sub> that are nearly homogeneous and that are kept in a high salt (0.6 M KCl), low glycerol (10%) buffer during the final purification step (as recommended in the published purification procedure), and then are stored in that buffer in the presence or absence of BSA at -70 °C, generally have reduced Int recombinase activity. But with all preparations tested, the activity can be increased by mixing Int with S20 before addition to an assay. We have found, however, that the activities of preparations of Int in the high salt buffer which appear lower can be increased to a certain extent by diluting the preparation in a low salt buffer (0.05 M KCl) before assay or more preferably by dialyzing the preparation into a buffer containing low salt (0.05 to 0.1 M KCl) and 50% (v/v) glycerol. Such preparations can then be stored at -20 °C or -70 °C. Furthermore, regardless of the level of recombinase activity these preparations have by themselves before or after dialysis, addition of appropriate amounts of S20 stimulates that activity.

20

25

30

### **Conclusions**

Taken together, these results demonstrate that at least two *E. coli* ribosomal proteins, S20 and L27, and possibly a third *E. coli* ribosomal protein, S15, stimulate  $\lambda$  Int-mediated recombination *in vitro*. In addition, purified preparations of  $\lambda$  Int that appear to be inactive in a  $\lambda$  recombination system can be restored to activity by the addition of S20.

**Example 2: Stimulation of Integrase Recombination by other *E. coli* Ribosomal Proteins**

In addition to S20 and L27, other *E. coli* ribosomal proteins may stimulate the activity of recombination systems, particularly the  $\lambda$  Int system. In particular, *E. coli* ribosomal proteins that are basic and are about 14 kilodaltons or less in size are used to stimulate the activity of prokaryotic recombination systems. Such ribosomal proteins that may be used are shown in Table 5:

5  
10  
15  
20  
25  
30  
TABLE 5: Additional Ribosomal Proteins for Use in Stimulating Recombination Activity

Ribosomal Protein	No. of Basic Residues (% of Total)	No. of Total Residues	Molecular Weight (Daltons)
S16	17 (16.5%)	103	11,736
S14	23 (23.7%)	97	11,063
S15	16 (18.4%)	87	10,001
S16	14 (17.1%)	82	9,191
S17	16 (19.3%)	83	9,573
S16	17 (23.0%)	74	8,896
S16	19 (20.9%)	91	10,299
S21	23 (32.9%)	70	8,369
L21	17 (16.5%)	103	11,565
L24	21 (21.2%)	99	11,013
L24	22 (21.4%)	103	11,185
L28	17 (18.1%)	83	10,694
L28	18 (23.4%)	77	8,875
L29	12 (19.0%)	63	7,274
L30	10 (17.2%)	58	6,411
L31	12 (19.4%)	52	6,971
L32	11 (19.6%)	54	6,315
L33	15 (27.8%)	54	6,255
L34	14 (30.4%)	46	5,381

These ribosomal proteins are isolated from natural sources as generally described above for S20 and L27 and as discussed in *Ann. Rev. Biochem* 51:155 (1982), *Ann. Rev. Biochem.* 52:35 (1983), *Ann. Rev. Biochem.* 53:75 (1984), and *Ann. Rev. Biochem.* 66:679 (1997). Alternatively, the ribosomal proteins are prepared by recombinant DNA methodologies as generally outlined above for the

5

10

production of S20 and Xis. Isolated ribosomal proteins are used to stimulate recombination activity, particularly that of Int, by including one or more of them in the compositions of the invention as described above for S20 and L27, and these compositions are used in integrative and excisive recombination assays, and in the recombinational cloning methods of the invention, as generally outlined in Example 1 for S20. In addition, ribosomal proteins corresponding to those described herein may be used in accordance with the invention. For example, ribosomal proteins from other prokaryotic sources, and from eukaryotic sources (e.g., yeast, fungi, animals (including mammals such as humans), plants, and the like) may be used in the methods and compositions of the invention.

6  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

Having now fully described the present invention in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious to one of ordinary skill in the art that the same can be performed by modifying or changing the invention within a wide and equivalent range of conditions, formulations and other parameters without affecting the scope of the invention or any specific embodiment thereof, and that such modifications or changes are intended to be encompassed within the scope of the appended claims.

All publications, patents and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains, and are herein incorporated by reference to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference.